



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/90, 15/85, C12Q 1/68, C12N 5/10, 9/12, 15/13, C07K 16/28, C12N 15/12, C07K 14/705, G01N 33/53, C12N 15/62, C07K 19/00		A1	(11) International Publication Number: WO 98/41645 (43) International Publication Date: 24 September 1998 (24.09.98)
(21) International Application Number: PCT/US98/03935 (22) International Filing Date: 9 March 1998 (09.03.98)		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
(30) Priority Data: 08/819,866 14 March 1997 (14.03.97) US 09/023,715 13 February 1998 (13.02.98) US		(Published) <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(71) Applicant: IDEC PHARMACEUTICALS CORPORATION [US/US]; 11011 Torreyana Road, San Diego, CA 92121 (US).			
(72) Inventors: REFF, Mitchell, E.; 4166 Combe Way, San Diego, CA 92122 (US). BARNETT, Richard, Spence; 306 Belmont Court, San Marcos, CA 92069 (US). McLACHLAN, Karen, Retta; Apartment B6, 766 South Nardo, Solana Beach, CA 92075 (US).			
(74) Agents: GESS, E., Joseph et al.; Burns, Doane, Swecker & Mathis L.L.P., P.O. Box 1404, Alexandria, VA 22313-1404 (US).			
(54) Title: METHOD FOR INTEGRATING GENES AT SPECIFIC SITES IN MAMMALIAN CELLS VIA HOMOLOGOUS RECOMBINATION AND VECTORS FOR ACCOMPLISHING THE SAME			
(57) Abstract A method for achieving site specific integration of a desired DNA at a target site in a mammalian cell via homologous recombination is described. This method provides for the reproducible selection of cell lines wherein a desired DNA is integrated at a predetermined transcriptionally active site previously marked with a marker plasmid. The method is particularly suitable for the production of mammalian cell lines which secrete mammalian proteins at high levels, in particular immunoglobulins. Vectors and vector combinations for use in the subject cloning method are also provided.			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		

Title of the Invention

METHOD FOR INTEGRATING GENES AT SPECIFIC SITES IN MAMMALIAN CELLS VIA HOMOLOGOUS RECOMBINATION AND VECTORS FOR ACCOMPLISHING THE SAME

5

Field of the Invention

The present invention relates to a process of targeting the integration of a desired exogenous DNA to a specific location within the genome of a mammalian cell.

10 More specifically, the invention describes a novel method for identifying a transcriptionally active target site ("hot spot") in the mammalian genome, and inserting a desired DNA at this site via homologous recombination.

15 The invention also optionally provides the ability for gene amplification of the desired DNA at this location by co-integrating an amplifiable selectable marker, e.g., DHFR, in combination with the exogenous DNA. The invention additionally describes the construction of novel vectors suitable for accomplishing the above, and

20 further provides mammalian cell lines produced by such methods which contain a desired exogenous DNA integrated at a target hot spot.

- 2 -

Background

Technology for expressing recombinant proteins in both prokaryotic and eukaryotic organisms is well established. Mammalian cells offer significant advantages over bacteria or yeast for protein production, resulting from their ability to correctly assemble, glycosylate and post-translationally modify recombinantly expressed proteins. After transfection into the host cells, recombinant expression constructs can be maintained as extrachromosomal elements, or may be integrated into the host cell genome. Generation of stably transfected mammalian cell lines usually involves the latter; a DNA construct encoding a gene of interest along with a drug resistance gene (dominant selectable marker) is introduced into the host cell, and subsequent growth in the presence of the drug allows for the selection of cells that have successfully integrated the exogenous DNA. In many instances, the gene of interest is linked to a drug resistant selectable marker which can later be subjected to gene amplification. The gene encoding dihydrofolate reductase (DHFR) is most commonly used for this purpose. Growth of cells in the presence of methotrexate, a competitive inhibitor of DHFR, leads to increased DHFR production by means of amplification of the DHFR gene. As flanking regions of DNA will also become amplified, the resultant coamplification of a DHFR linked gene in the transfected cell line can lead to increased protein

production, thereby resulting in high level expression of the gene of interest.

While this approach has proven successful, there are a number of problems with the system because of the random nature of the integration event. These problems exist because expression levels are greatly influenced by the effects of the local genetic environment at the gene locus, a phenomena well documented in the literature and generally referred to as "position effects" (for example, see Al-Shawi et al, *Mol. Cell. Biol.*, 10:1192-1198 (1990); Yoshimura et al, *Mol. Cell. Biol.*, 7:1296-1299 (1987)). As the vast majority of mammalian DNA is in a transcriptionally inactive state, random integration methods offer no control over the transcriptional fate of the integrated DNA. Consequently, wide variations in the expression level of integrated genes can occur, depending on the site of integration. For example, integration of exogenous DNA into inactive, or transcriptionally "silent" regions of the genome will result in little or no expression. By contrast integration into a transcriptionally active site may result in high expression.

Therefore, when the goal of the work is to obtain a high level of gene expression, as is typically the desired outcome of genetic engineering methods, it is generally necessary to screen large numbers of transfec-tants to find such a high producing clone.

Additionally, random integration of exogenous DNA into the genome can in some instances disrupt important cellular genes, resulting in an altered phenotype. These factors can make the generation of high expressing stable mammalian cell lines a complicated and laborious process.

Recently, our laboratory has described the use of DNA vectors containing translationally impaired dominant selectable markers in mammalian gene expression. (This 10 is disclosed in U.S. Serial No. 08/147,696 filed November 3, 1993, recently allowed).

These vectors contain a translationally impaired neomycin phosphotransferase (neo) gene as the dominant selectable marker, artificially engineered to contain an 15 intron into which a DHFR gene along with a gene or genes of interest is inserted. Use of these vectors as expression constructs has been found to significantly reduce the total number of drug resistant colonies produced, thereby facilitating the screening procedure in 20 relation to conventional mammalian expression vectors.

Furthermore, a significant percentage of the clones obtained using this system are high expressing clones. These results are apparently attributable to the modifications made to the neo selectable marker. Due to 25 the translational impairment of the neo gene, transfected cells will not produce enough neo protein to survive drug selection, thereby decreasing the overall

number of drug resistant colonies. Additionally, a higher percentage of the surviving clones will contain the expression vector integrated into sites in the genome where basal transcription levels are high, resulting in overproduction of neo, thereby allowing the cells to overcome the impairment of the neo gene. Concomitantly, the genes of interest linked to neo will be subject to similar elevated levels of transcription. This same advantage is also true as a result of the artificial intron created within neo; survival is dependent on the synthesis of a functional neo gene, which is in turn dependent on correct and efficient splicing of the neo introns. Moreover, these criteria are more likely to be met if the vector DNA has integrated into a region which is already highly transcriptionally active.

Following integration of the vector into a transcriptionally active region, gene amplification is performed by selection for the DHFR gene. Using this system, it has been possible to obtain clones selected using low levels of methotrexate (50nM), containing few (<10) copies of the vector which secrete high levels of protein (>55pg/cell/day). Furthermore, this can be achieved in a relatively short period of time. However, the success in amplification is variable. Some transcriptionally active sites cannot be amplified and

- 6 -

therefore the frequency and extent of amplification from a particular site is not predictable.

Overall, the use of these translationally impaired vectors represents a significant improvement over other 5 methods of random integration. However, as discussed, the problem of lack of control over the integration site remains a significant concern.

One approach to overcome the problems of random integration is by means of gene targeting, whereby the 10 exogenous DNA is directed to a specific locus within the host genome. The exogenous DNA is inserted by means of homologous recombination occurring between sequences of DNA in the expression vector and the corresponding homologous sequence in the genome. However, while this 15 type of recombination occurs at a high frequency naturally in yeast and other fungal organisms, in higher eukaryotic organisms it is an extremely rare event. In mammalian cells, the frequency of homologous versus non-homologous (random integration) recombination is reported to range from 1/100 to 1/5000 (for example, see 20 Capecchi, *Science*, 244:1288-1292 (1989); Morrow and Kucherlapati, *Curr. Op. Biotech.*, 4:577-582 (1993)).

One of the earliest reports describing homologous recombination in mammalian cells comprised an artificial 25 system created in mouse fibroblasts (Thomas et al, *Cell*, 44:419-428 (1986)). A cell line containing a mutated, non-functional version of the neo gene integrated into

- 7 -

the host genome was created, and subsequently targeted with a second non-functional copy of neo containing a different mutation. Reconstruction of a functional neo gene could occur only by gene targeting. Homologous 5 recombinants were identified by selecting for G418 resistant cells, and confirmed by analysis of genomic DNA isolated from the resistant clones.

Recently, the use of homologous recombination to replace the heavy and light immunoglobulin genes at 10 endogenous loci in antibody secreting cells has been reported. (U.S. Patent No. 5,202,238, Fell et al, 1993.) However, this particular approach is not widely applicable, because it is limited to the production of immunoglobulins in cells which 15 endogenously express immunoglobulins, e.g., B cells and myeloma cells. Also, expression is limited to single copy gene levels because co-amplification after homologous recombination is not included. The method is further complicated by the fact that two separate 20 integration events are required to produce a functional immunoglobulin: one for the light chain gene followed by one for the heavy chain gene.

An additional example of this type of system has 25 been reported in NS/0 cells, where recombinant immunoglobulins are expressed by homologous recombination into the immunoglobulin gamma 2A locus (Hollis et al, international patent application #

- 8 -

PCT/IB95 (00014).) Expression levels obtained from this site were extremely high - on the order of 20pg/cell/day from a single copy integrant. However, as in the above example, expression is limited to this level because an 5 amplifiable gene is not cointegrated in this system. Also, other researchers have reported aberrant glycosylation of recombinant proteins expressed in NS/0 cells (for example, see Flesher et al, Biotech. and Bioeng., 48:399-407 (1995)), thereby limiting the 10 applicability of this approach.

The cre-loxP recombination system from bacteriophage P1 has recently been adapted and used as a means of gene targeting in eukaryotic cells. Specifically, the site specific integration of exogenous 15 DNA into the Chinese hamster ovary (CHO) cell genome using cre recombinase and a series of lox containing vectors have been described. (Fukushige and Sauer, Proc. Natl. Acad. Sci. USA, 89:7905-7909 (1992).) This system is attractive in that it provides for 20 reproducible expression at the same chromosomal location. However, no effort was made to identify a chromosomal site from which gene expression is optimal, and as in the above example, expression is limited to single copy levels in this system. Also, it is 25 complicated by the fact that one needs to provide for expression of a functional recombinase enzyme in the mammalian cell.

The use of homologous recombination between an introduced DNA sequence and its endogenous chromosomal locus has also been reported to provide a useful means of genetic manipulation in mammalian cells, as well as 5 in yeast cells. (See e.g., Bradley et al, *Meth. Enzymol.*, 223:855-879 (1993); Capecchi, *Science*, 244:1288-1292 (1989); Rothstein et al, *Meth. Enzymol.*, 194:281-301 (1991)). To date, most mammalian gene targeting studies have been directed toward gene 10 disruption ("knockout") or site-specific mutagenesis of selected target gene loci in mouse embryonic stem (ES) cells. The creation of these "knockout" mouse models has enabled scientists to examine specific structure-function issues and examine the biological 15 importance of a myriad of mouse genes. This field of research also has important implications in terms of potential gene therapy applications.

Also, vectors have recently been reported by Cell-tech (Kent, U.K.) which purportedly are targeted to 20 transcriptionally active sites in NSO cells, which do not require gene amplification (Peakman et al, *Hum. Antibod. Hybridomas*, 5:65-74 (1994)). However, levels of immunoglobulin secretion in these unamplified cells have not been reported to exceed 20pg/cell/day, while in 25 amplified CHO cells, levels as high as 100pg/cell/day can be obtained (Id.).

- 10 -

It would be highly desirable to develop a gene targeting system which reproducibly provided for the integration of exogenous DNA into a predetermined site in the genome known to be transcriptionally active.

5 Also, it would be desirable if such a gene targeting system would further facilitate co-amplification of the inserted DNA after integration. The design of such a system would allow for the reproducible and high level expression of any cloned gene of interest in a mammalian 10 cell, and undoubtedly would be of significant interest to many researchers.

In this application, we provide a novel mammalian expression system, based on homologous recombination occurring between two artificial substrates contained in 15 two different vectors. Specifically, this system uses a combination of two novel mammalian expression vectors, referred to as a "marking" vector and a "targeting" vector.

Essentially, the marking vector enables the identification and marking of a site in the mammalian genome which is transcriptionally active, i.e., a site at which gene expression levels are high. This site can be regarded as a "hot spot" in the genome. After integration of the marking vector, the subject expression system 25 enables another DNA to be integrated at this site, i.e., the targeting vector, by means of homologous recombination occurring between DNA sequences common to

- 11 -

both vectors. This system affords significant advantages over other homologous recombination systems.

Unlike most other homologous systems employed in mammalian cells, this system exhibits no background.

5 Therefore, cells which have only undergone random integration of the vector do not survive the selection.

Thus, any gene of interest cloned into the targeting plasmid is expressed at high levels from the marked hot spot. Accordingly, the subject method of gene expression substantially or completely eliminates the problems inherent to systems of random integration, discussed in detail above. Moreover, this system provides reproducible and high level expression of any recombinant protein at the same transcriptionally active site in the mammalian genome. In addition, gene amplification may be effected at this particular transcriptionally active site by including an amplifiable dominant selectable marker (e.g. DHFR) as part of the marking vector.

Objects of the Invention

20 Thus, it is an object of the invention to provide an improved method for targeting a desired DNA to a specific site in a mammalian cell.

It is a more specific object of the invention to provide a novel method for targeting a desired DNA to a specific site in a mammalian cell via homologous recombination.

- 12 -

It is another specific object of the invention to provide novel vectors for achieving site specific integration of a desired DNA in a mammalian cell.

5 It is still another object of the invention to provide novel mammalian cell lines which contain a desired DNA integrated at a predetermined site which provides for high expression.

10 It is a more specific object of the invention to provide a novel method for achieving site specific integration of a desired DNA in a Chinese hamster ovary (CHO) cell.

15 It is another more specific object of the invention to provide a novel method for integrating immunoglobulin genes, or any other genes, in mammalian cells at predetermined chromosomal sites that provide for high expression.

20 It is another specific object of the invention to provide novel vectors and vector combinations suitable for integrating immunoglobulin genes into mammalian cells at predetermined sites that provide for high expression.

25 It is another object of the invention to provide mammalian cell lines which contain immunoglobulin genes integrated at predetermined sites that provide for high expression.

It is an even more specific object of the invention to provide a novel method for integrating immunoglobulin

- 13 -

genes into CHO cells that provide for high expression, as well as novel vectors and vector combinations that provide for such integration of immunoglobulin genes into CHO cells.

5 In addition, it is a specific object of the invention to provide novel CHO cell lines which contain immunoglobulin genes integrated at predetermined sites that provide for high expression, and have been amplified by methotrexate selection to secrete even greater amounts
10 of functional immunoglobulins.

Brief Description of the Figures

Figure 1 depicts a map of a marking plasmid according to the invention referred to as Desmond. The plasmid is shown in circular form (1a) as well as a
15 linearized version used for transfection (1b).

Figure 2(a) shows a map of a targeting plasmid referred to "Molly". Molly is shown here encoding the anti-CD20 immunoglobulin genes, expression of which is described in Example 1.

20 Figure 2(b) shows a linearized version of Molly, after digestion with the restriction enzymes *Kpn*1 and *Pac*1. This linearized form was used for transfection.

Figure 3 depicts the potential alignment between Desmond sequences integrated into the CHO genome, and
25 incoming targeting Molly sequences. One potential ar-

- 14 -

rangement of Molly integrated into Desmond after homologous recombination is also presented.

Figure 4 shows a Southern analysis of single copy Desmond clones. Samples are as follows:

- 5 Lane 1: λ HindIII DNA size marker
- Lane 2: Desmond clone 10F3
- Lane 3: Desmond clone 10C12
- Lane 4: Desmond clone 15C9
- Lane 5: Desmond clone 14B5
- 10 Lane 6: Desmond clone 9B2

Figure 5 shows a Northern analysis of single copy Desmond clones. Samples are as follows: Panel A: northern probed with CAD and DHFR probes, as indicated on the figure. Panel B: duplicate northern, probed with 15 CAD and HisD probes, as indicated. The RNA samples loaded in panels A and B are as follows:

- Lane 1: clone 9B2, lane 2; clone 10C12, lane 3; clone 14B5, lane 4; clone 15C9, lane 5; control RNA from CHO transfected with a HisD and DHFR containing plasmid,
- 20 lane 6; untransfected CHO.

Figure 6 shows a Southern analysis of clones resulting from the homologous integration of Molly into Desmond. Samples are as follows:

- Lane 1: λ HindIII DNA size markers, Lane 2: 20F4, lane 3; 25
- 5F9, lane 4; 21C7, lane 5; 24G2, lane 6; 25E1, lane 7; 28C9, lane 8; 29F9, lane 9; 39G11, lane 10; 42F9, lane 11; 50G10, lane 12; Molly plasmid DNA, linearized with

- 15 -

BglII (top band) and cut with BglII and KpnI (lower band), lane 13; untransfected Desmond.

Figures 7A through 7G contain the Sequence Listing for Desmond.

5 Figures 8A through 8I contain the Sequence Listing for Molly-containing anti-CD20.

Figure 9 contains a map of the targeting plasmid, "Mandy," shown here encoding anti-CD23 genes, the expression of which is disclosed in Example 5.

10 Figures 10A through 10N contain the sequence listing of "Mandy" containing the anti-CD23 genes as disclosed in Example 5.

Detailed Description of the Invention

The invention provides a novel method for integrating a desired exogenous DNA at a target site within the genome of a mammalian cell via homologous recombination. 15 Also, the invention provides novel vectors for achieving the site specific integration of a DNA at a target site in the genome of a mammalian cell.

20 More specifically, the subject cloning method provides for site specific integration of a desired DNA in a mammalian cell by transfection of such cell with a "marker plasmid" which contains a unique sequence that is foreign to the mammalian cell genome and which 25 provides a substrate for homologous recombination, followed by transfection with a "target plasmid" containing

a sequence which provides for homologous recombination with the unique sequence contained in the marker plasmid, and further comprising a desired DNA that is to be integrated into the mammalian cell. Typically, the 5 integrated DNA will encode a protein of interest, such as an immunoglobulin or other secreted mammalian glycoprotein.

The exemplified homologous recombination system uses the neomycin phosphotransferase gene as a dominant 10 selectable marker. This particular marker was utilized based on the following previously published observations;

(i) the demonstrated ability to target and restore function to a mutated version of the neo gene (cited 15 earlier) and

(ii) our development of translationally impaired expression vectors, in which the neo gene has been artificially created as two exons with a gene of interest inserted in the intervening intron; neo exons are correctly spliced and translated *in vivo*, producing a functional protein and thereby conferring G418 resistance on the resultant cell population. In this application, the neo gene is split into three exons. The third exon of neo is present on the "marker" plasmid and becomes integrated into the host cell genome upon integration of the marker plasmid into the mammalian cells. Exons 1 and 2 are present on the targeting plasmid, and are separated 20 25

by an intervening intron into which at least one gene of interest is cloned. Homologous recombination of the targeting vector with the integrated marking vector results in correct splicing of all three exons of the 5 neo gene and thereby expression of a functional neo protein (as determined by selection for G418 resistant colonies). Prior to designing the current expression system, we had experimentally tested the functionality of such a triply spliced neo construct in mammalian 10 cells. The results of this control experiment indicated that all three neo exons were properly spliced and therefore suggested the feasibility of the subject invention.

However, while the present invention is exemplified 15 using the neo gene, and more specifically a triple split neo gene, the general methodology should be efficacious with other dominant selectable markers.

As discussed in greater detail *infra*, the present invention affords numerous advantages to conventional 20 gene expression methods, including both random integration and gene targeting methods. Specifically, the subject invention provides a method which reproducibly allows for site-specific integration of a desired DNA into a transcriptionally active domain of a mammalian 25 cell. Moreover, because the subject method introduces an artificial region of "homology" which acts as a unique substrate for homologous recombination and the

insertion of a desired DNA, the efficacy of subject invention does not require that the cell endogenously contain or express a specific DNA. Thus, the method is generically applicable to all mammalian cells, and can
5 be used to express any type of recombinant protein.

The use of a triply spliced selectable marker, e.g., the exemplified triply spliced neo construct, guarantees that all G418 resistant colonies produced will arise from a homologous recombination event (random 10 integrants will not produce a functional neo gene and consequently will not survive G418 selection). Thus, the subject invention makes it easy to screen for the desired homologous event. Furthermore, the frequency of additional random integrations in a cell that has undergone a homologous recombination event appears to be low.
15

Based on the foregoing, it is apparent that a significant advantage of the invention is that it substantially reduces the number of colonies that need be screened to identify high producer clones, i.e., cell 20 lines containing a desired DNA which secrete the corresponding protein at high levels. On average, clones containing integrated desired DNA may be identified by screening about 5 to 20 colonies (compared to several thousand which must be screened when using standard 25 random integration techniques, or several hundred using the previously described intronic insertion vectors) Additionally, as the site of integration was preselected

- 19 -

and comprises a transcriptionally active domain, all exogenous DNA expressed at this site should produce comparable, i.e. high levels of the protein of interest.

Moreover, the subject invention is further advantageous in that it enables an amplifiable gene to be inserted on integration of the marking vector. Thus, when a desired gene is targeted to this site via homologous recombination, the subject invention allows for expression of the gene to be further enhanced by gene amplification. In this regard, it has been reported in from the literature that different genomic sites have different capacities for gene amplification (Meinkoth et al, *Mol. Cell Biol.*, 7:1415-1424 (1987)). Therefore, this technique is further advantageous as it allows for the placement of a desired gene of interest at a specific site that is both transcriptionally active and easily amplified. Therefore, this should significantly reduce the amount of time required to isolate such high producers.

Specifically, while conventional methods for the construction of high expressing mammalian cell lines can take 6 to 9 months, the present invention allows for such clones to be isolated on average after only about 3-6 months. This is due to the fact that conventionally isolated clones typically must be subjected to at least three rounds of drug resistant gene amplification in order to reach satisfactory levels of gene expression.

- 20 -

As the homologously produced clones are generated from a preselected site which is a high expression site, fewer rounds of amplification should be required before reaching a satisfactory level of production.

5 Still further, the subject invention enables the reproducible selection of high producer clones wherein the vector is integrated at low copy number, typically single copy. This is advantageous as it enhances the stability of the clones and avoids other potential adverse side-effects associated with high copy number. As described *supra*, the subject homologous recombination system uses the combination of a "marker plasmid" and a "targeting plasmid" which are described in more detail below.

10

15 The "marker plasmid" which is used to mark and identify a transcriptionally hot spot will comprise at least the following sequences:

(i) a region of DNA that is heterologous or unique to the genome of the mammalian cell, which functions as
20 a source of homology, allows for homologous recombination (with a DNA contained in a second target plasmid). More specifically, the unique region of DNA (i) will generally comprise a bacterial, viral, yeast synthetic, or other DNA which is not normally present in the
25 mammalian cell genome and which further does not comprise significant homology or sequence identity to DNA contained in the genome of the mammalian cell.

Essentially, this sequence should be sufficiently different to mammalian DNA that it will not significantly recombine with the host cell genome via homologous recombination. The size of such unique DNA 5 will generally be at least about 2 to 10 kilobases in size, or higher, more preferably at least about 10kb, as several other investigators have noted an increased frequency of targeted recombination as the size of the homology region is increased (Capecchi, *Science*, 10 244:1288-1292 (1989)).

The upper size limit of the unique DNA which acts as a site for homologous recombination with a sequence in the second target vector is largely dictated by potential stability constraints (if DNA is too large it 15 may not be easily integrated into a chromosome and the difficulties in working with very large DNAs.

(ii) a DNA including a fragment of a selectable marker DNA, typically an exon of a dominant selectable marker gene. The only essential feature of this DNA is 20 that it not encode a functional selectable marker protein unless it is expressed in association with a sequence contained in the target plasmid. Typically, the target plasmid will comprise the remaining exons of the dominant selectable marker gene (those not comprised in 25 "targeting" plasmid). Essentially, a functional selectable marker should only be produced if homologous recombination occurs (resulting in the association and

expression of this marker DNA (i) sequence together with the portion(s) of the selectable marker DNA fragment which is (are) contained in the target plasmid).

As noted, the current invention exemplifies the
5 use of the neomycin phosphotransferase gene as the dominant selectable marker which is "split" in the two vectors. However, other selectable markers should also be suitable, e.g., the *Salmonella histidinol dehydrogenase* gene, *hygromycin phosphotransferase* gene, *herpes simplex* 10 virus *thymidine kinase* gene, *adenosine deaminase* gene, *glutamine synthetase* gene and *hypoxanthine-guanine phosphoribosyl transferase* gene.

(iii) a DNA which encodes a functional selectable marker protein, which selectable marker is different
15 from the selectable marker DNA (ii). This selectable marker provides for the successful selection of mammalian cells wherein the marker plasmid is successfully integrated into the cellular DNA. More preferably, it is desirable that the marker plasmid comprise two such
20 dominant selectable marker DNAs, situated at opposite ends of the vector. This is advantageous as it enables integrants to be selected using different selection agents and further enables cells which contain the entire vector to be selected. Additionally, one marker
25 can be an amplifiable marker to facilitate gene amplification as discussed previously. Any of the

dominant selectable marker listed in (ii) can be used as well as others generally known in the art.

Moreover, the marker plasmid may optionally further comprise a rare endonuclease restriction site. This is potentially desirable as this may facilitate cleavage.

If present, such rare restriction site should be situated close to the middle of the unique region that acts as a substrate for homologous recombination. Preferably such sequence will be at least about 12 nucleotides.

10 The introduction of a double stranded break by similar methodology has been reported to enhance the frequency of homologous recombination. (Choulika et al, *Mol. Cell. Biol.*, 15:1968-1973 (1995)). However, the presence of such sequence is not essential.

15 The "targeting plasmid" will comprise at least the following sequences:

(1) the same unique region of DNA contained in the marker plasmid or one having sufficient homology or sequence identity therewith that said DNA is capable of combining via homologous recombination with the unique region (i) in the marker plasmid. Suitable types of DNAs are described *supra* in the description of the unique region of DNA (1) in the marker plasmid.

20
25 (2) The remaining exons of the dominant selectable marker, one exon of which is included as (ii) in the marker plasmid listed above. The essential features of this DNA fragment is that it result in a functional

(selectable) marker protein only if the target plasmid integrates via homologous recombination (wherein such recombination results in the association of this DNA with the other fragment of the selectable marker DNA contained in the marker plasmid) and further that it allow for insertion of a desired exogenous DNA. Typically, this DNA will comprise the remaining exons of the selectable marker DNA which are separated by an intron. For example, this DNA may comprise the first two exons of the neo gene and the marker plasmid may comprise the third exon (back third of neo).

(3) The target plasmid will also comprise a desired DNA, e.g., one encoding a desired polypeptide, preferably inserted within the selectable marker DNA fragment contained in the plasmid. Typically, the DNA will be inserted in an intron which is comprised between the exons of the selectable marker DNA. This ensures that the desired DNA is also integrated if homologous recombination of the target plasmid and the marker plasmid occurs. This intron may be naturally occurring or it may be engineered into the dominant selectable marker DNA fragment.

This DNA will encode any desired protein, preferably one having pharmaceutical or other desirable properties. Most typically the DNA will encode a mammalian protein, and in the current examples provided, an immunoglobulin or an immunoadhesin. However the

invention is not in any way limited to the production of immunoglobulins.

As discussed previously, the subject cloning method is suitable for any mammalian cell as it does not require for efficacy that any specific mammalian sequence or sequences be present. In general, such mammalian cells will comprise those typically used for protein expression, e.g., CHO cells, myeloma cells, COS cells, BHK cells, Sp2/0 cells, NIH 3T3 and HeLa cells. In the examples which follow, CHO cells were utilized. The advantages thereof include the availability of suitable growth medium, their ability to grow efficiently and to high density in culture, and their ability to express mammalian proteins such as immunoglobulins in biologically active form.

Further, CHO cells were selected in large part because of previous usage of such cells by the inventors for the expression of immunoglobulins (using the translationally impaired dominant selectable marker containing vectors described previously). Thus, the present laboratory has considerable experience in using such cells for expression. However, based on the examples which follow, it is reasonable to expect similar results will be obtained with other mammalian cells.

In general, transformation or transfection of mammalian cells according to the subject invention will be effected according to conventional methods. So that the

invention may be better understood, the construction of exemplary vectors and their usage in producing integrants is described in the examples below.

EXAMPLE 1

5

Design and Preparation of Marker
and Targeting Plasmid DNA Vectors

The marker plasmid herein referred to as "Desmond" was assembled from the following DNA elements:

(a) Murine dihydrofolate reductase gene (DHFR),
10 incorporated into a transcription cassette, comprising the mouse beta globin promoter 5" to the DHFR start site, and bovine growth hormone poly adenylation signal 3" to the stop codon. The DHFR transcriptional cassette was isolated from TCAE6, an expression vector created
15 previously in this laboratory (Newman et al, 1992, *Bio-technology*, 10:1455-1460).

(b) E. coli β-galactosidase gene - commercially available, obtained from Promega as pSV-b-galactosidase control vector, catalog # E1081.

20 (c) Baculovirus DNA, commercially available, purchased from Clontech as pBAKPAK8, cat # 6145-1.

(d) Cassette comprising promoter and enhancer elements from Cytomegalovirus and SV40 virus. The cassette was generated by PCR using a derivative of expression
25 vector TCAE8 (Reff et al, *Blood*, 83:435-445 (1994)). The enhancer cassette was inserted within the baculo-

virus sequence, which was first modified by the insertion of a multiple cloning site.

(e) E. coli GUS (glucuronidase) gene, commercially available, purchased from Clontech as pB101, cat. # 5 6017-1.

(f) Firefly luciferase gene, commercially available, obtained from Promega as pGEM-Luc (catalog # E1541).

(g) S. typhimurium histidinol dehydrogenase gene 10 (HisD). This gene was originally a gift from (Donahue et al, Gene, 18:47-59 (1982)), and has subsequently been incorporated into a transcription cassette comprising the mouse beta globin major promoter 5' to the gene, and the SV40 polyadenylation signal 3' to the gene.

15 The DNA elements described in (a)-(g) were combined into a pBR derived plasmid backbone to produce a 7.7kb contiguous stretch of DNA referred to in the attached figures as "homology". Homology in this sense refers to sequences of DNA which are not part of the mammalian genome and are used to promote homologous recombination 20 between transfected plasmids sharing the same homology DNA sequences.

(h) Neomycin phosphotransferase gene from TNS (Davis and Smith, Ann. Rev. Micro., 32:469-518 (1978)).

25 The complete neo gene was subcloned into pBluescript SK- (Stratagene catalog # 212205) to facilitate genetic manipulation. A synthetic linker was then inserted into

a unique PstI site occurring across the codons for amino acid 51 and 52 of neo. This linker encoded the necessary DNA elements to create an artificial splice donor site, intervening intron and splice acceptor site within the neo gene, thus creating two separate exons, presently referred to as neo exon 1 and 2. Neo exon 1 encodes the first 51 amino acids of neo, while exon 2 encodes the remaining 203 amino acids plus the stop codon of the protein A NotI cloning site was also created within the 10 intron.

Neo exon 2 was further subdivided to produce neo exons 2 and 3. This was achieved as follows: A set of PCR primers were designed to amplify a region of DNA encoding neo exon 1, intron and the first 111 2/3 amino acids of exon2. The 3' PCR primer resulted in the introduction of a new 5' splice site immediately after the second nucleotide of the codon for amino acid 111 in exon 2, therefore generating a new smaller exon 2. The DNA fragment now encoding the original exon 1, intron and new exon 2 was then subcloned and propagated in a pBR based vector. The remainder of the original exon 2 was used as a template for another round of PCR amplification, which generated "exon3". The 5' primer for this round of amplification introduced a new splice acceptor site at the 5' side of the newly created exon 25 3, i.e. before the final nucleotide of the codon for amino acid 111. The resultant 3 exons of neo encode the

following information: exon 1 - the first 51 amino acids of neo; exon 2 - the next 111 2/3 amino acids, and exon 3 the final 91 1/3 amino acids plus the translational stop codon of the neo gene.

5 Neo exon 3 was incorporated along with the above mentioned DNA elements into the marking plasmid "Desmond". Neo exons 1 and 2 were incorporated into the targeting plasmid "Molly". The Not1 cloning site created within the intron between exons 1 and 2 was used in
10 subsequent cloning steps to insert genes of interest into the targeting plasmid.

A second targeting plasmid "Mandy" was also generated. This plasmid is almost identical to "Molly" (some restriction sites on the vector have been changed) except that the original HisD and DHFR genes contained in "Molly" were inactivated. These changes were incorporated because the Desmond cell line was no longer being cultured in the presence of Histidinol, therefore it seemed unnecessary to include a second copy of the
15 HisD gene. Additionally, the DHFR gene was inactivated to ensure that only a single DHFR gene, namely the one present in the Desmond marked site, would be amplifiable in any resulting cell lines. "Mandy" was derived from "Molly" by the following modifications:
20

25 (i) A synthetic linker was inserted in the middle of the DHFR coding region. This linker created a stop codon and shifted the remainder of the DHFR coding

- 30 -

region out of frame, therefore rendering the gene nonfunctional.

(ii) A portion of the HisD gene was deleted and replaced with a PCR generated HisD fragment lacking the promoter and start codon of the gene.

Figure 1 depicts the arrangement of these DNA elements in the marker plasmid "Desmond". Figure 2 depicts the arrangement of these elements in the first targeting plasmid, "Molly". Figure 3 illustrates the possible arrangement in the CHO genome, of the various DNA elements after targeting and integration of Molly DNA into Desmond marked CHO cells. Figure 9 depicts the targeting plasmid "Mandy."

Construction of the marking and targeting plasmids from the above listed DNA elements was carried out following conventional cloning techniques (see, e.g., Molecular Cloning, A Laboratory Manual, J. Sambrook et al, 1987, Cold Spring Harbor Laboratory Press, and Current Protocols in Molecular Biology, F. M. Ausubel et al, eds., 1987, John Wiley and Sons). All plasmids were propagated and maintained in E. coli XLI blue (Stratagene, cat. # 200236). Large scale plasmid preparations were prepared using Promega Wizard Maxiprep DNA Purification System®, according to the manufacturer's directions.

- 31 -

EXAMPLE 2

Construction of a Marked CHO Cell Line

1. Cell Culture and Transfection Procedures to
Produced Marked CHO Cell Line

5 Marker plasmid DNA was linearized by digestion overnight at 37°C with Bst1107I. Linearized vector was ethanol precipitated and resuspended in sterile TE to a concentration of 1mg/ml. Linearized vector was introduced into DHFR-Chinese hamster ovary cells (CHO cells)
10 DG44 cells (Urlaub et al, *Som. Cell and Mol. Gen.*, 12:555-566 (1986)) by electroporation as follows...

Exponentially growing cells were harvested by centrifugation, washed once in ice cold SBS (sucrose buffered solution, 272mM sucrose, 7mM sodium phosphate, pH 7.4, 1mM magnesium chloride) then resuspended in SBS to a concentration of 10⁷ cells/ml. After a 15 minute incubation on ice, 0.4ml of the cell suspension was mixed with 40µg linearized DNA in a disposable electroporation cuvette. Cells were shocked using a BTX electrocell manipulator (San Diego, CA) set at 230 volts, 400 microfaraday capacitance, 13 ohm resistance. Shocked cells were then mixed with 20 ml of prewarmed CHO growth media (CHO-S-SFMII, Gibco/BRL, catalog # 31033-012) and plated in 96 well tissue culture plates.
20 Forty eight hours after electroporation, plates were fed with selection media (in the case of transfection with Desmond, selection media is CHO-S-SFMII without

- 32 -

hypoxanthine or thymidine, supplemented with 2mM Histidinol (Sigma catalog # H6647)). Plates were maintained in selection media for up to 30 days, or until some of the wells exhibited cell growth. These cells 5 were then removed from the 96 well plates and expanded ultimately to 120 ml spinner flasks where they were maintained in selection media at all times.

EXAMPLE 3

Characterization of Marked CHO Cell Lines

10 (a) Southern Analysis

Genomic DNA was isolated from all stably growing Desmond marked CHO cells. DNA was isolated using the Invitrogen Easy® DNA kit, according to the manufacturer's directions. Genomic DNA was then digested with 15 HindIII overnight at 37°C, and subjected to Southern analysis using a PCR generated digoxigenin labelled probe specific to the DHFR gene. Hybridizations and washes were carried out using Boehringer Mannheim's DIG easy hyb (catalog # 1603 558) and DIG Wash and Block 20 Buffer Set (catalog # 1585 762) according to the manufacturer's directions. DNA samples containing a single band hybridizing to the DHFR probe were assumed to be Desmond clones arising from a single cell which had integrated a single copy of the plasmid. These clones 25 were retained for further analysis. Out of a total of 45 HisD resistant cell lines isolated, only 5 were

single copy integrants. Figure 4 shows a Southern blot containing all 5 of these single copy Desmond clones. Clone names are provided in the figure legend.

(b) Northern Analysis

5 Total RNA was isolated from all single copy Desmond clones using TRIzol reagent (Gibco/BRL cat # 15596-026) according to the manufacturer's directions. 10-20 μ g RNA from each clone was analyzed on duplicate formaldehyde gels. The resulting blots were probed with PCR generated digoxigenin labelled DNA probes to (i) DHFR message, (ii) HisD message and (iii) CAD message. CAD is a trifunctional protein involved in uridine biosynthesis (Wahl et al, *J. Biol. Chem.*, 254, 17:8679-8689 (1979)), and is expressed equally in all cell types. It is used here as an internal control to help quantitate RNA loading. Hybridizations and washes were carried out using the above mentioned Boehringer Mannheim reagents. The results of the Northern analysis are shown in Figure 5. The single copy Desmond clone exhibiting the highest levels of both the His D and DHFR message is clone 15C9, shown in lane 4 in both panels of the figure. This clone was designated as the "marked cell line" and used in future targeting experiments in CHO, examples of which are presented in the following sections.

EXAMPLE 4Expression of Anti-CD20 Antibody
in Desmond Marked CHO Cells

C2B8, a chimeric antibody which recognizes B-cell surface antigen CD20, has been cloned and expressed previously in our laboratory. (Reff et al, *Blood*, 83:434-45 (1994)). A 4.1 kb DNA fragment comprising the C2B8 light and heavy chain genes, along with the necessary regulatory elements (eukaryotic promoter and polyadenylation signals) was inserted into the artificial intron created between exons 1 and 2 of the neo gene contained in a pBR derived cloning vector. This newly generated 5kb DNA fragment (comprising neo exon 1, C2B8 and neo exon 2) was excised and used to assemble the targeting plasmid Molly. The other DNA elements used in the construction of Molly are identical to those used to construct the marking plasmid Desmond, identified previously. A complete map of Molly is shown in Fig. 2.

The targeting vector Molly was linearized prior to transfection by digestion with *Kpn*1 and *Pac*1, ethanol precipitated and resuspended in sterile TE to a concentration of 1.5mg/mL. Linearized plasmid was introduced into exponentially growing Desmond marked cells essentially as described, except that 80 μ g DNA was used in each electroporation. Forty eight hours postelectroporation, 96 well plates were supplemented with selection medium - CHO-SSFMII supplemented with 400 μ g/mL Geneti-

- 35 -

cin (G418, Gibco/BRL catalog # 10131-019). Plates were maintained in selection medium for up to 30 days, or until cell growth occurred in some of the wells. Such growth was assumed to be the result of clonal expansion of a single G418 resistant cell. The supernatants from all G418 resistant wells were assayed for C2B8 production by standard ELISA techniques, and all productive clones were eventually expanded to 120mL spinner flasks and further analyzed.

10 Characterization of Antibody secreting Targeted Cells

A total of 50 electroporations with Molly targeting plasmid were carried out in this experiment, each of which was plated into separate 96 well plates. A total of 10 viable, anti-CD20 antibody secreting clones were obtained and expanded to 120ml spinner flasks. Genomic DNA was isolated from all clones, and Southern analyses were subsequently performed to determine whether the clones represented single homologous recombination events or whether additional random integrations had occurred in the same cells. The methods for DNA isolation and Southern hybridization were as described in the previous section. Genomic DNA was digested with EcoRI and probed with a PCR generated digoxigenin labelled probe to a segment of the CD20 heavy chain constant region. The results of this Southern analysis are presented in figure 6. As can be seen in the figure, 8 of

- 36 -

the 10 clones show a single band hybridizing to the CD20 probe, indicating a single homologous recombination event has occurred in these cells. Two of the ten, clones 24G2 and 28C9, show the presence of additional 5 band(s), indicative of an additional random integration elsewhere in the genome.

We examined the expression levels of anti-CD20 antibody in all ten of these clones, the data for which is shown in Table 1, below.

10

Table 1:

Expression Level of Anti-CD20
Secreting Homologous Integrants

	<u>Clone</u>	<u>Anti-CD20, pg/c/d</u>
15	20F4	3.5
	25E1	2.4
	42F9	1.8
	39G11	1.5
	21C7	1.3
	50G10	0.9
20	29F9	0.8
	5F9	0.3
<hr/>		
	28C9*	4.5
	24G2*	2.1

5

* These clones contained additional randomly integrated copies of anti-CD20. Expression levels of these clones therefore reflect a contribution from both the homologous and random sites.

Expression levels are reported as picogram per cell per day (pg/c/d) secreted by the individual clones, and represented the mean levels obtained from three separate ELISAs on samples taken from 120 mL spinner flasks.

10

As can be seen from the data, there is a variation in antibody secretion of approximately ten fold between the highest and lowest clones. This was somewhat unexpected as we anticipated similar expression levels from all clones due to the fact the anti-CD20 genes are all integrated into the same Desmond marked site. Nevertheless, this observed range in expression extremely small in comparison to that seen using any traditional random integration method or with our translationally impaired vector system.

15

Clone 20F4, the highest producing single copy integrant was selected for further study. Table 2 (below) presents ELISA and cell culture data from seven day production runs of this clone.

- 38 -

Table 2:

7 Day Production Run Data for 20F4

Day	% Viable	Viable/ml (x 10 ⁵)	T _{x2} (hr)	mg/L	pg/c/d
1	96	3.4	31	1.3	4.9
5	2	6	29	2.5	3.4
3	94	9.9	33	4.7	3.2
4	90	17.4	30	6.8	3
5	73	14		8.3	
6	17	3.5		9.5	

10 Clone 20F4 was seeded at 2x10⁵ml in a 120ml spinner flask on day 0. On the following six days, cell counts were taken, doubling times calculated and 1ml samples of supernatant removed from the flask and analyzed for secreted anti-CD20 by ELISA.

15 This clone is secreting on average, 3-5pg antibody/- cell/day, based on this ELISA data. This is the same level as obtained from other high expressing single copy clones obtained previously in our laboratory using the previously developed translationally impaired random 20 integration vectors. This result indicates the following:

25 (1) that the site in the CHO genome marked by the Desmond marking vector is highly transcriptionally active, and therefore represents an excellent site from which to express recombinant proteins, and

- 39 -

(2) that targeting by means of homologous recombination can be accomplished using the subject vectors and occurs at a frequency high enough to make this system a viable and desirable alternative to random integration methods.

To further demonstrate the efficacy of this system, we have also demonstrated that this site is amplifiable, resulting in even higher levels of gene expression and protein secretion. Amplification was achieved by plating serial dilutions of 20F4 cells, starting at a density of 2.5×10^4 cells/ml, in 96 well tissue culture dishes, and culturing these cells in media (CHO-SSFMII) supplemented with 5, 10, 15 or 20nM methotrexate. Antibody secreting clones were screened using standard ELISA techniques, and the highest producing clones were expanded and further analyzed. A summary of this amplification experiment is presented in Table 3 below.

- 40 -

Table 3:

Summary of 20F4 Amplification

nM MTX	# Wells Assayed	Expression Level mg/l 96 well	# Wells Expanded	Expression Level pg/c/d from spinner
10	56	3-13	4	10-15
5	27	2-14	3	15-18
20	17	4-11	1	ND

Methotrexate amplification of 20F4 was set up as described in the text, using the concentrations of methotrexate indicated in the above table. Supernatants from all surviving 96 well colonies were assayed by ELISA, and the range of anti-CD20 expressed by these clones is indicated in column 3. Based on these results, the highest producing clones were expanded to 120ml spinners and several ELISAs conducted on the spinner supernatants to determine the pg/cell/day expression levels, reported in column 5.

The data here clearly demonstrates that this site can be amplified in the presence of methotrexate. Clones from the 10 and 15nM amplifications were found to produce on the order of 15-20pg/cell/day.

A 15nM clone, designated 20F4-15A5, was selected as the highest expressing cell line. This clone originated from a 96 well plate in which only 22 wells grew, and was therefore assumed to have arisen from a single cell. A 15nM clone, designated 20F4-15A5, was selected as the highest expressing cell line. This clone originated

- 41 -

from a 96 well plate in which only 22 wells grew, and was therefore assumed to have arisen from a single cell. The clone was then subjected to a further round of methotrexate amplification. As described above, serial 5 dilutions of the culture were plated into 96 well dishes and cultured in CHO-SS-FMII medium supplemented with 200, 300 or 400nM methotrexate. Surviving clones were screened by ELISA, and several high producing clones were expanded to spinner cultures and further analyzed.

10 A summary of this second amplification experiment is presented in Table 4.

Table 4:
Summary of 20F4-15A5 Amplification

nM MTX	# Wells Assayed	Expression Level mg/l 96 well	# Wells Expanded	Expression Level pg/c/d, spinner
15 200	67	23-70	1	50-60
250	86	21-70	4	55-60
300	81	15-75	3	40-50

20 Methotrexate amplifications of 20F4-15A5 were set up and assayed as described in the text. The highest producing wells, the numbers of which are indicated in column 4, were expanded to 120ml spinner flasks. The expression levels of the cell lines derived from these wells is recorded as pg/c/d in column 5.

25 The highest producing clone came from the 250nM methotrexate amplification. The 250nM clone, 20F4-15A5-250A6 originated from a 96 well plate in which only wells

- 42 -

grew, and therefore is assumed to have arisen from a single cell. Taken together, the data in Tables 3 and 4 strongly indicates that two rounds of methotrexate amplification are sufficient to reach expression levels of 5 60pg/cell/day, which is approaching the maximum secretion capacity of immunoglobulin in mammalian cells (Reff, M.E., *Curr. Opin. Biotech.*, 4:573-576 (1993)). The ability to reach this secretion capacity with just two amplification steps further enhances the utility of 10 this homologous recombination system. Typically, random integration methods require more than two amplification steps to reach this expression level and are generally less reliable in terms of the ease of amplification. Thus, the homologous system offers a more efficient and 15 time saving method of achieving high level gene expression in mammalian cells.

EXAMPLE 5

Expression of Anti-Human CD23 Antibody
in Desmond Marked CHO Cells

20 CD23 is low affinity IgE receptor which mediates binding of IgE to B and T lymphocytes (Sutton, B.J., and Gould, H.J., *Nature*, 366:421-428 (1993)). Anti-human CD23 monoclonal antibody 5E8 is a human gamma-1 monoclonal antibody recently cloned and expressed in our 25 laboratory. This antibody is disclosed in commonly

- 43 -

assigned Serial No. 08/803,085, filed on February 20, 1997.

The heavy and light chain genes of 5E8 were cloned into the mammalian expression vector N5KG1, a derivative 5 of the vector NEOSPLA (Barnett et al., in *Antibody Expression and Engineering*, H.Y Yang and T. Imanaka, eds., pp27-40 (1995)) and two modifications were then made to the genes. We have recently observed somewhat higher secretion of immunoglobulin light chains compared to 10 heavy chains in other expression constructs in the laboratory (Reff et al, 1997, unpublished observations). In an attempt to compensate for this deficit, we altered the 5E8 heavy chain gene by the addition of a stronger promoter/enhancer element immediately upstream of the 15 start site. In subsequent steps, a 2.9kb DNA fragment comprising the 5E8 modified light and heavy chain genes was isolated from the N5KG1 vector and inserted into the targeting vector Mandy. Preparation of 5E8-containing Molly and electroporation into Desmond 15C9 CHO cells 20 was essentially as described in the preceding section.

One modification to the previously described protocol was in the type of culture medium used. Desmond marked CHO cells were cultured in protein-free CD-CHO medium (Gibco-BRL, catalog # AS21206) supplemented with 25 3mg/L recombinant insulin (3mg/mL stock, Gibco-BRL, catalog # AS22057) and 8mM L-glutamine (200mM stock, Gibco-BRL, catalog # 25030-081). Subsequently, trans-

fected cells were selected in the above medium supplemented with 400 μ g/mL geneticin. In this experiment, 20 electroporations were performed and plated into 96 well tissue culture dishes. Cells grew and secreted anti-

5 CD23 in a total of 68 wells, all of which were assumed to be clones originating from a single G418 cell.

Twelve of these wells were expanded to 120ml spinner flasks for further analysis. We believe the increased number of clones isolated in this experiment (68 com-

10 pared with 10 for anti-CD20 as described in Example 4) is due to a higher cloning efficiency and survival rate of cells grown in CD-CHO medium compared with CHO-SS-FMII medium. Expression levels for those clones analyzed in spinner culture ranged from 0.5-3pg/c/d, in

15 close agreement with the levels seen for the anti-CD20 clones. The highest producing anti-CD23 clone, designated 4H12, was subjected to methotrexate amplification in order to increase its expression levels. This amplification was set up in a manner similar to that described

20 for the anti-CD20 clone in Example 4. Serial dilutions of exponentially growing 4H12 cells were plated into 96 well tissue culture dishes and grown in CD-CHO medium supplemented with 3mg/L insulin, 8mM glutamine and 30, 35 or 40nM methotrexate. A summary of this

25 amplification experiment is presented in Table 5.

Table 5:

- 45 -

Summary of 2H12 Amplification

nM MTX	# Wells Assayed	Expression Level mg/l 96 well	# Wells Expanded	Expression Level pg/c/d from spinner
30	100	6-24	8	10-25
35	64	4-27	2	10-15
5	40	4-20	1	ND

The highest expressing clone obtained was a 30nM clone, isolated from a plate on which 22 wells had grown. This clone, designated 4H12-30G5, was reproducibly secreting 18-22pg antibody per cell per day. This is the same range of expression seen for the first amplification of the anti CD20 clone 20F4 (clone 20F4-15A5 which produced 15-18pg/c/d, as described in Example 4). This data serves to further support the observation that amplification at this marked site in CHO is reproducible and efficient. A second amplification of this 30nM cell line is currently underway. It is anticipated that saturation levels of expression will be achievable for the anti-CD23 antibody in just two amplification steps, as was the case for anti-CD20.

20

EXAMPLE 6

Expression of Immunoadhesin in Desmond Marked CHO Cells

CTLA-4, a member of the Ig superfamily, is found on the surface of T lymphocytes and is thought to play a role in antigen-specific T-cell activation (Dariavach et al, *Eur. J. Immunol.*, 18:1901-1905 (1988); and Linsley et al, *J. Exp. Med.*, 174:561-569 (1991)). In order to further study the precise role of the CTLA-4 molecule in the activation pathway, a soluble fusion protein comprising the extracellular domain of CTLA-4 linked to a truncated form of the human IgG1 constant region was

- 46 -

created (Linsley et al (Id.). We have recently expressed this CTLA-4 Ig fusion protein in the mammalian expression vector BLECH1, a derivative of the plasmid NEOSPLA (Barnett et al, in Antibody Expression and Engineering, H.Y Yang and T. Imanaka, eds., pp27-40 (1995)). An 800bp fragment encoding the CTLA-4 Ig was isolated from this vector and inserted between the SacII and BglII sites in Molly.

Preparation of CTLA-4Ig-Molly and electroporation into Desmond clone 15C9 CHO cells was performed as described in the previous example relating to anti-CD20. Twenty electroporations were carried out, and plated into 96 well culture dishes as described previously. Eighteen CTLA-4 expressing wells were isolated from the 96 well plates and carried forward to the 120ml spinner stage. Southern analyses on genomic DNA isolated from each of these clones were then carried out to determine how many of the homologous clones contained additional random integrants. Genomic DNA was digested with BglII and probed with a PCR generated digoxigenin labelled probe to the human IgG1 constant region. The results of this analysis indicated that 85% of the CTLA-4 clones are homologous integrants only; the remaining 15% contained one additional random integrant. This result corroborates the findings from the expression of anti-CD20 discussed above, where 80% of the clones were single homologous integrants. Therefore, we can conclude

that this expression system reproducibly yields single targeted homologous integrants in at least 80% of all clones produced.

Expression levels for the homologous CTLA4-Ig clones ranged from 8-12 pg/cell/day. This is somewhat higher than the range reported for anti-CD20 antibody and anti-CD23 antibody clones discussed above. However, we have previously observed that expression of this molecule using the intronic insertion vector system also resulted in significantly higher expression levels than are obtained for immunoglobulins. We are currently unable to provide an explanation for this observation.

EXAMPLE 7

Targeting Anti-CD20 to an alternate Desmond Marked CHO Cell Line

As we described in a preceding section, we obtained 5 single copy Desmond marked CHO cell lines (see Figures 4 and 5). In order to demonstrate that the success of our targeting strategy is not due to some unique property of Desmond clone 15C9 and limited only to this clone, we introduced anti-CD20 Molly into Desmond clone 9B2 (lane 6 in figure 4, lane 1 in figure 5). Preparation of Molly DNA and electroporation into Desmond 9B2 was exactly as described in the previous example pertaining to anti-CD20. We obtained one homologous integrant from this experiment. This clone was expanded to a 120ml

spinner flask, where it produced on average 1.2pg anti-CD20/cell/day. This is considerably lower expression than we observed with Molly targeted into Desmond 15C9. However, this was the anticipated result, based on our 5 northern analysis of the Desmond clones. As can be seen in Figure 5, mRNA levels from clone 9B2 are considerably lower than those from 15C9, indicating the site in this clone is not as transcriptionally active as that in 15C9. Therefore, this experiment not only demonstrates 10 the reproducibility of the system - presumably any marked Desmond site can be targeted with Molly - it also confirms the northern data that the site in Desmond 15C9 is the most transcriptionally active.

From the foregoing, it will be appreciated that, 15 although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without diverting from the scope of the invention. Accordingly, the invention is not limited by the appended claims.

WHAT IS CLAIMED IS:

1. A method for inserting a desired DNA at a target site in the genome of a mammalian cell which comprises the following steps:

5 (i) transfected or transforming a mammalian cell with a first plasmid ("marker plasmid") containing the following sequences:

10 (a) a region of DNA that is heterologous to the mammalian cell genome which when integrated in the mammalian cell genome provides a unique site for homologous recombination;

(b) a DNA fragment encoding a portion of a first selectable marker protein; and

15 (c) at least one other selectable marker DNA that provides for selection of mammalian cells which have been successfully integrated with the marker plasmid;

(ii) selecting a cell which contain the marker plasmid integrated in its genome;

20 (iii) transfected or transforming said selected cell with a second plasmid ("target plasmid") which contains the following sequences:

25 (a) a region of DNA that is identical or is sufficiently homologous to the unique region in the marker plasmid such that this region of DNA can recombine with said DNA via homologous recombination;

- 50 -

(b) a DNA fragment encoding a portion of the same selectable marker contained in the marker plasmid, wherein the active selectable marker protein encoded by said DNA is only produced if said fragment is expressed 5 in association with the fragment of said selectable marker DNA contained in the marker plasmid; and

(iv) selecting cells which contain the target plasmid integrated at the target site by screening for the expression of the first selectable marker protein.

10 2. The method of Claim 1, wherein the DNA fragment encoding a fragment of a first selectable marker is an exon of a dominant selectable marker.

15 3. The method of Claim 2, wherein the second plasmid contains the remaining exons of said first selectable marker.

4. The method of Claim 3, wherein at least one DNA encoding a desired protein is inserted between said exons of said first selectable marker contained in the target plasmid.

20 5. The method of Claim 4, wherein a DNA encoding a dominant selectable marker is further inserted between the exons of said first selectable marker contained in

the target plasmid to provide for co-amplification of the DNA encoding the desired protein.

6. The method of Claim 3, wherein the first dominant selectable marker is selected from the group consisting of neomycin phosphotransferase, histidinol dehydrogenase, dihydrofolate reductase, hygromycin phosphotransferase, herpes simplex virus thymidine kinase, adenosine deaminase, glutamine synthetase, and hypoxanthine-guanine phosphoribosyl transferase.

10 7. The method of Claim 4, wherein the desired protein is a mammalian protein.

8. The method of Claim 7, wherein the protein is an immunoglobulin.

15 9. The method of Claim 1, which further comprises determining the RNA levels of the selectable marker (c) contained in the marker plasmid prior to integration of the target vector.

20 10. The method of Claim 9, wherein the other selectable marker contained in the marker plasmid is a dominant selectable marker selected from the group consisting of histidinol dehydrogenase, herpes simplex

- 52 -

thymidine kinase, hydromycin phosphotransferase, adenosine deaminase and glutamine synthetase.

11. The method of Claim 1, wherein the mammalian cell is selected from the group consisting of Chinese hamster ovary (CHO) cells, myeloma cells, baby hamster kidney cells, COS cells, NSO cells, HeLa cells and NIH 3T3 cells.

12. The method of Claim 11, wherein the cell is a CHO cell.

10 13. The method of Claim 1, wherein the marker plasmid contains the third exon of the neomycin phosphotransferase gene and the target plasmid contains the first two exons of the neomycin phosphotransferase gene.

15 14. The method of Claim 1, wherein the marker plasmid further contains a rare restriction endonuclease sequence which is inserted within the region of homology.

20 15. The method of Claim 1, wherein the unique region of DNA that provides for homologous recombination is a bacterial DNA, a viral DNA or a synthetic DNA.

- 53 -

16. The method of Claim 1, wherein the unique region of DNA that provides for homologous recombination is at least 300 nucleotides.

17. The method of Claim 16, wherein the unique 5 region of DNA ranges in size from about 300 nucleotides to 20 kilobases.

18. The method of claim 17, wherein the unique region of DNA preferably ranges in size from 2 to 10 kilobases.

10 19. The method of Claim 1, wherein the first selectable marker DNA is split into at least three exons.

20. The method of Claim 1, wherein the unique region of DNA that provides for homologous recombination 15 is a bacterial DNA, an insect DNA, a viral DNA or a synthetic DNA.

21. The method of Claim 20, wherein the unique region of DNA does not contain any functional genes.

20 22. A vector system for inserting a desired DNA at a target site in the genome of a mammalian cell which comprises at least the following:

(i) a first plasmid ("marker plasmid") containing at least the following sequences:

(a) a region of DNA that is heterologous to the mammalian cell genome which when integrated in the mammalian cell genome provides a unique site for homologous recombination;

(b) a DNA fragment encoding a portion of a first selectable marker protein; and

(c) at least one other selectable marker DNA that provides for selection of mammalian cells which have been successfully integrated with the marker plasmid; and

(ii) a second plasmid ("target plasmid") which contains at least the following sequences:

(a) a region of DNA that is identical or is sufficiently homologous to the unique region in the marker plasmid such that this region of DNA can recombine with said DNA via homologous recombination;

(b) a DNA fragment encoding a portion of the same selectable marker contained in the marker plasmid, wherein the active selectable marker protein encoded by said DNA is only produced if said fragment is expressed in association with the fragment of said selectable marker DNA contained in the marker plasmid.

- 55 -

23. The vector system of Claim 22, wherein the DNA fragment encoding a fragment of a first selectable marker is an exon of a dominant selectable marker.

24. The vector system of Claim 23, wherein the
5 second plasmid contains the remaining exons of said
first selectable marker.

25. The vector system of Claim 24, wherein at least one DNA encoding a desired protein is inserted between said exons of said first selectable marker contained in the target plasmid.
10

26. The vector system of Claim 24, wherein a DNA encoding a dominant selectable marker is further inserted between the exons of said first selectable marker contained in the target plasmid to provide for co-amplification of the DNA encoding the desired protein.
15

27. The vector system of Claim 24, wherein the first dominant selectable marker is selected from the group consisting of neomycin phosphotransferase, histidinol dehydrogenase, dihydrofolate reductase, hygromycin phosphotransferase, herpes simplex virus thymidine kinase, adenosine deaminase, glutamine synthetase, and hypoxanthine-guanine phosphoribosyl transferase.
20

- 56 -

28. The vector system of Claim 25, wherein the desired protein is a mammalian protein.

29. The vector system of Claim 28, wherein the protein is an immunoglobulin.

5 30. The vector system of Claim 22, wherein the other selectable marker contained in the marker plasmid is a dominant selectable marker selected from the group consisting of histidinol dehydrogenase, herpes simplex thymidine kinase, hydromycin phosphotransferase, adenosine deaminase and glutamine synthetase.

10 31. The vector system of Claim 22, which provides for insertion of a desired DNA at a targeted site in the genome of a mammalian cell selected from the group consisting of Chinese hamster ovary (CHO) cells, myeloma cells, baby hamster kidney cells, COS cells, NSO cells, HeLa cells and NIH 3T3 cells.

15 32. The vector system of Claim 31, wherein the mammalian cell is a CHO cell.

20 33. The vector system of Claim 22, wherein the marker plasmid contains the third exon of the neomycin phosphotransferase gene and the target plasmid contains

- 57 -

the first two exons of the neomycin phosphotransferase gene.

34. The vector system of Claim 22, wherein the marker plasmid further contains a rare restriction endonuclease sequence which is inserted within the region of homology.

5 35. The vector system of Claim 22, wherein the unique region of DNA that provides for homologous recombination is a bacterial DNA, a viral DNA or a synthetic
10 DNA.

36. The vector system of Claim 22, wherein the unique region of DNA (a) contained in the marker plasmid vector system that provides for homologous recombination is at least 300 nucleotides.

15 37. The vector system of Claim 36, wherein the unique region of DNA ranges in size from about 300 nucleotides to 20 kilobases.

20 38. The vector system of Claim 37, wherein the unique region of DNA preferably ranges in size from 2 to 10 kilobases.

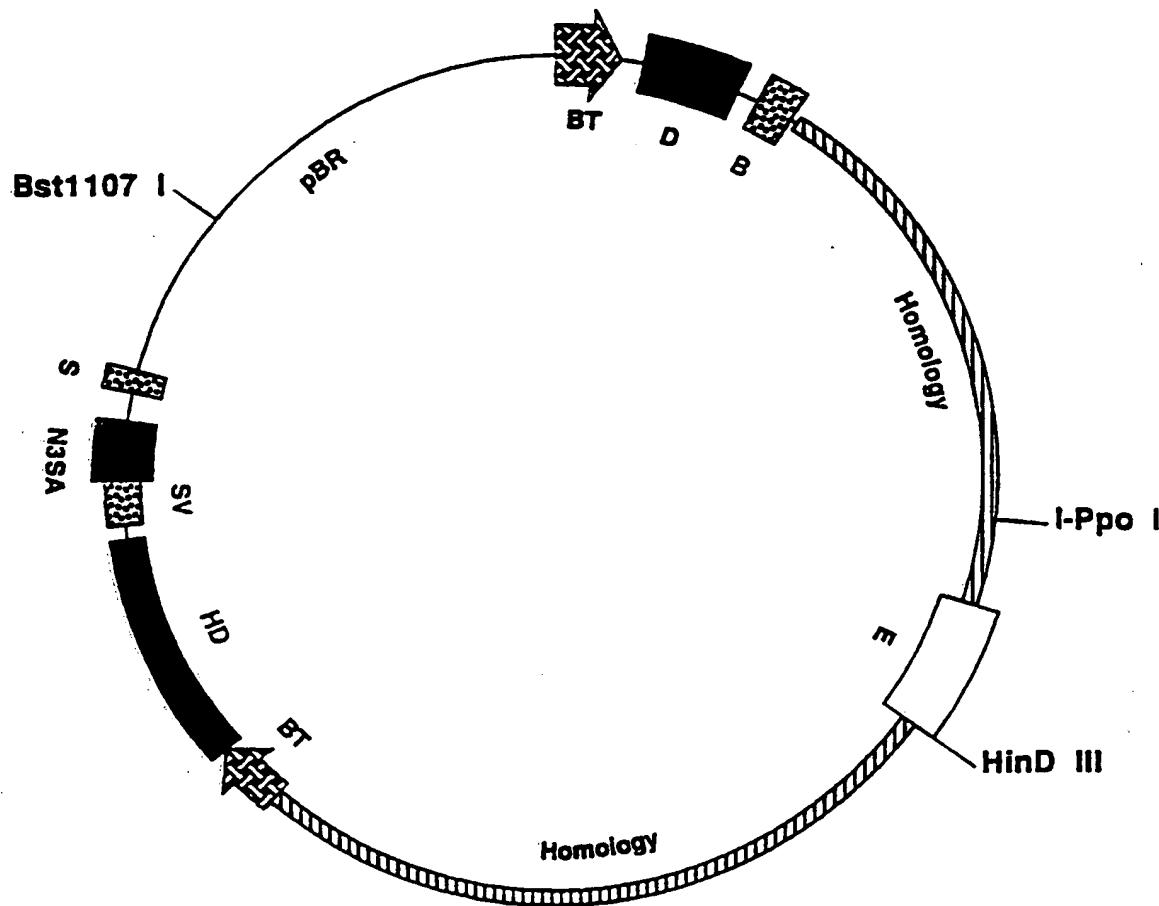
- 58 -

39. The vector system of Claim 22, wherein the first selectable marker DNA is split into at least three exons.

40. The vector system of Claim 22, wherein the
5 unique region of DNA that provides for homologous recombination is a bacterial DNA, an insect DNA, a viral DNA or a synthetic DNA.

41. The vector system of Claim 40, wherein the
unique region of DNA does not contain any functional
10 genes.

DESMOND



HD = Salmonella HisD Gene

N3 = Neomycin Phosphotransferase Exon 3

D = Murine Dihydrofolate reductase

E = Cytomegalovirus and SV40 Enhancers

SA = Splice acceptor

BT = Mouse Beta Globin Major Promoter

B = Bovine Growth Hormone Polyadenylation

S = SV40 Early Polyadenylation

SV = SV40 Late Polyadenylation

FIGURE 1A

Desmond

14,683 bp Bst1107 I linear

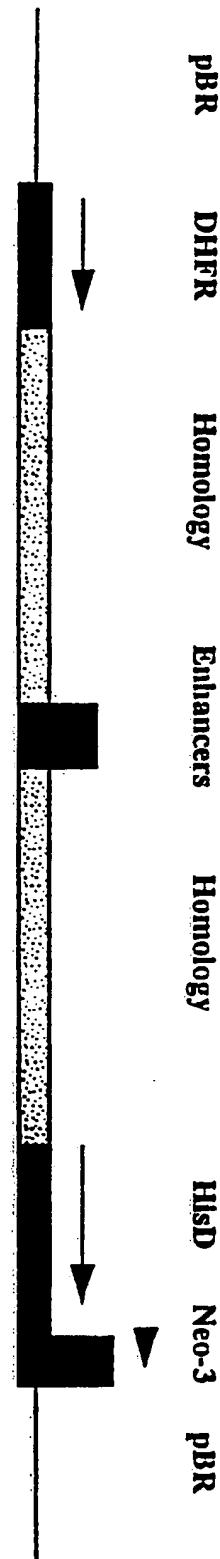
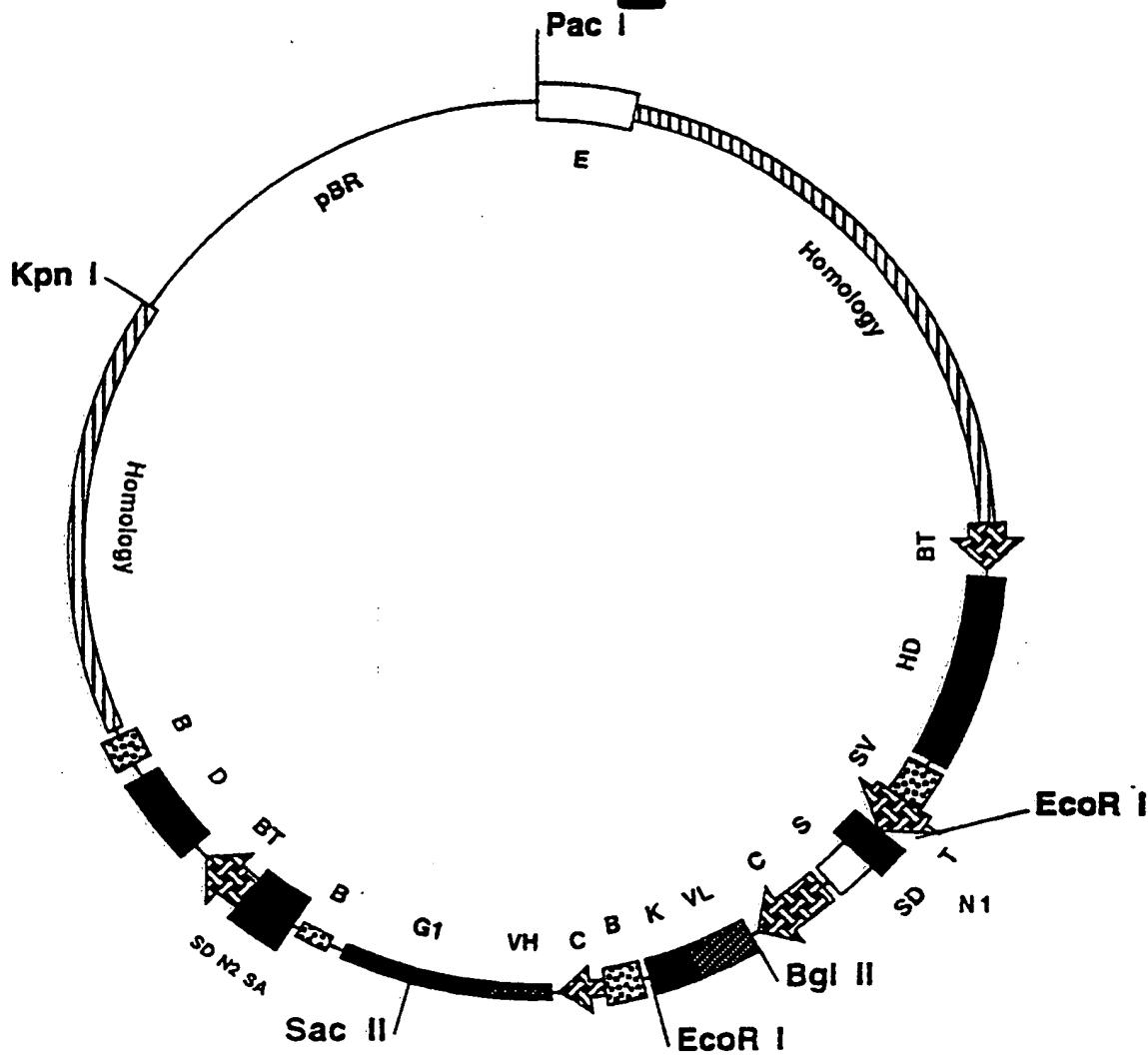


FIGURE 1B

Molly



D = Dihydrofolate reductase

N1 = Neomycin Phosphotransferase Exon 1

N2 = Neomycin Phosphotransferase Exon 2

VL = Anti-CD20 Light chain leader + Variable

K = Human Kappa Constant

VH = Anti-CD20 Heavy chain Leader + Variable

G1 = Human Gamma 1 Constant

HD = Salmonella Histidinol Dehydrogenase

E = CMV and SV40 enhancers S = SV40 Origin

SD = Splice donor SA = Splice acceptor

C = CMV promoter/enhancer

T = HSV TK promoter and Polyoma enhancers

BT = Mouse Beta Globin Major Promoter

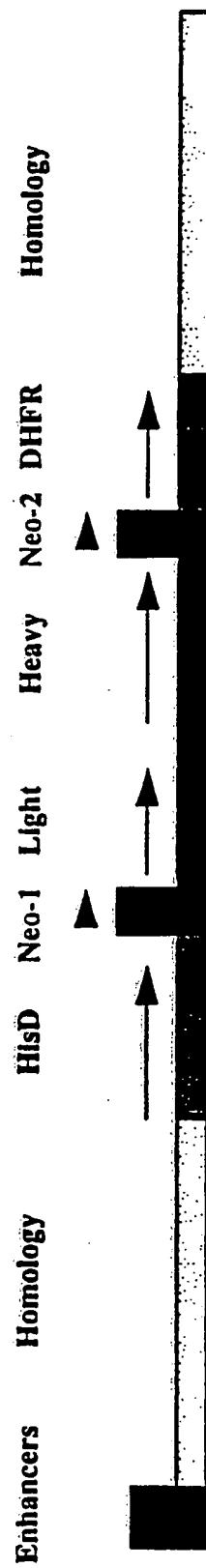
SV = SV40 Late Polyadenylation

B = Bovine Growth Hormone Polyadenylation

FIGURE 2A

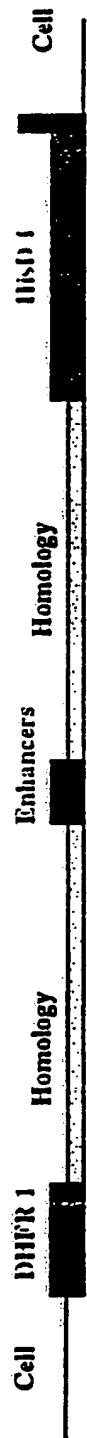
Molly

15,987 bp Pac I, Kpn I fragment

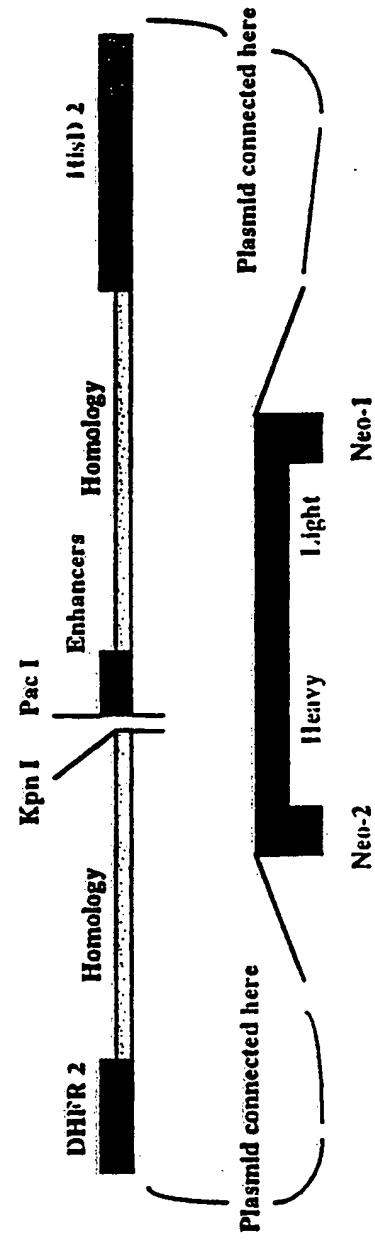


Homologous Recombination

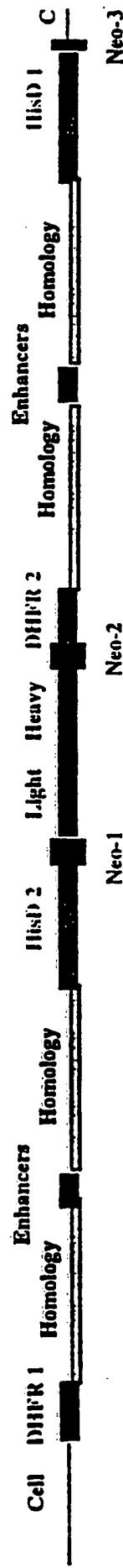
Desmond in CHO



Molly



Single crossover in CHO



Southern Analysis of Desmond Marked CHO Cells

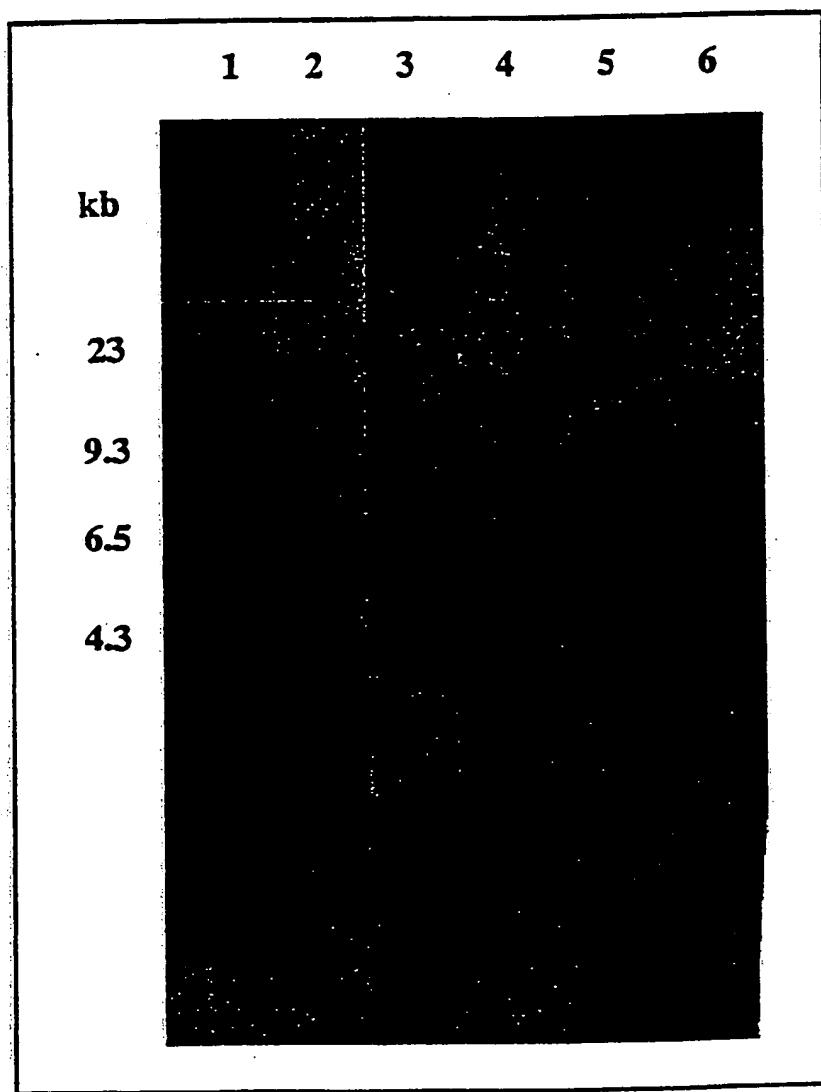


FIGURE 4

Northern Analysis of Desmond
Marked CHO Cells

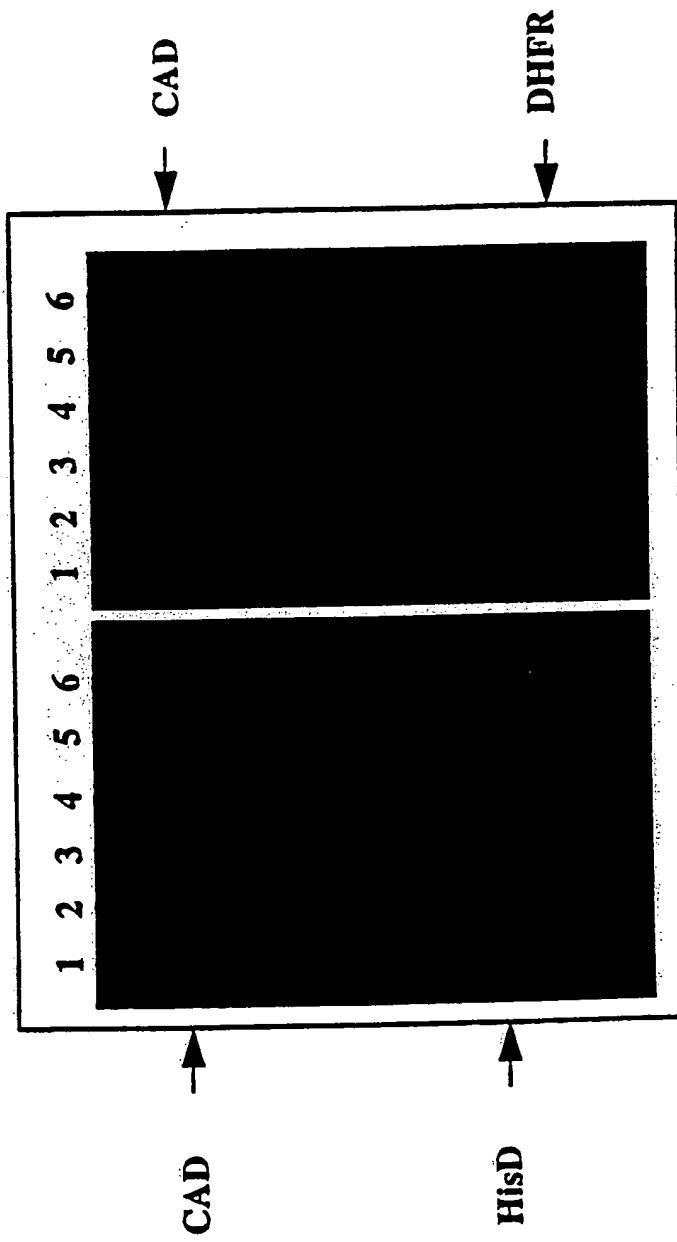


FIGURE 5

Southern Analysis of Anti CD20 Integrants in Marked CHO Cells

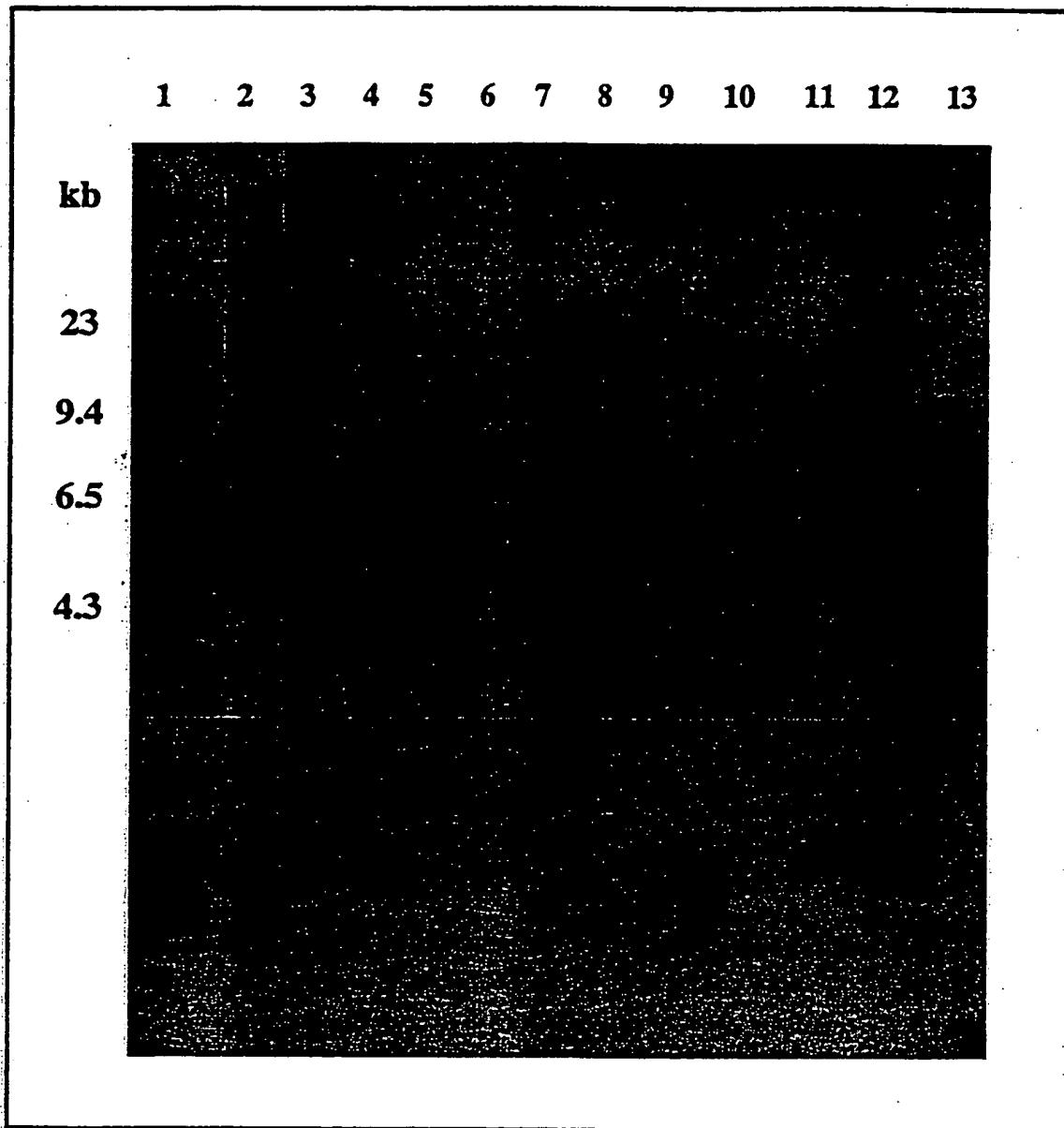


FIGURE 6

DNASIS
Desmond

10 20 30 40 50 60
 TTTCTAGACC TAGGGCGGCC AGCTAGTAGC TTTGCTTCTC AATTCTTAT TTGCATAATG
 70 80 90 100 110 120
 AGAAAAAAAG GAAAATTAAT TTTAACACCA ATTCAAGTAGT TGATTGAGCA AATGCCTTC
 130 140 150 160 170 180
 CAAAAAGGAT GCTTTAGAGA CAGTGTTCTC TGACAGATA AGGACAAACA TTATTCAGAG
 190 200 210 220 230 240
 GGAGTACCCA GAGCTGAGAC TCCTAACCCA GTGAGTGGCA CAGCATCCAG GGAGAAATAT
 250 260 270 280 290 300
 GCTTGTATC ACCGAAGCCT GATTCCGTAG AGCCACACCC TGTTAAGGGC CAATCTGCTC
 310 320 330 340 350 360
 ACACAGGATA GAGAGGGCAG GAGCCAGGGC AGAGCATATA AGGTGAGGTA GGATCAGTTG
 370 380 390 400 410 420
 CTCACAT TTGCTTCTGA CATAGTTGTG TTGGGAGCTT GGATAGCTTG GGGGGGGGAC
 430 440 450 460 470 480
 AGCTCAGGGC TGCGATTTCG CGCCAAACTT GACGGCAATC CTAGCGTGAA GGCTGGTAGG
 490 500 510 520 530 540
 ATTTTATCCC CGCTGCCATC ATGGTTCGAC CATTGAAC TG CATCGTCGCC GTGTCCAAA
 550 560 570 580 590 600
 ATATGGGGAT TGGCAAGAAC GGAGACCTAC CCTGGCCTCC GCTCAGGAAC GAGTTCAAGT
 610 620 630 640 650 660
 ACTTCCAAAG AATGACCACA ACCTCTTCAG TGGAAGGTA ACAGAATCTG GTGATTATGG
 670 680 690 700 710 720
 GTAGGAAAC CTGGTTCTCC ATTCTGAGA AGAATCGACC TTAAAGGAC AGAATTAATA
 730 740 750 760 770 780
 TTCTCAG TAGAGAACTC AAAGAACAC CACGAGGAGC TCATTTCTT GCCAAAAGTT
 790 800 810 820 830 840
 TGGATGATGC CTTAACACTT ATTGAACAAAC CGGAATTGGC AAGTAAAGTA GACATGGTTT
 850 860 870 880 890 900
 GGATAGTCGG AGGCAGTTCT GTTTACCAAG AAGCCATGAA TCAACCAGGC CACCTCAGAC
 910 920 930 940 950 960
 TCTTGTGAC AAGGATCATG CAGGAATTG AAAGTGACAC GTTTTCCCA GAAATTGATT
 970 980 990 1000 1010 1020
 TGGGGAAATA TAAACTTCTC CCAGAACACC CAGGCGTCCT CTCTGAGGTC CAGGAGGAAA
 1030 1040 1050 1060 1070 1080
 AAGGCATCAA GTATAAGTTT GAAGTCTACG AGAAGAAAGA CTAACAGGAA GATGCTTCA
 1090 1100 1110 1120 1130 1140
 AGTTCTCTGC TCCCCTCCCTA AAGCTATGCA TTTTATAAG ACCATGGGAC TTTTGCTGGC
 1150 1160 1170 1180 1190 1200
 TTAGATCAG CCTCGACTGT GCCTTCTAGT TGCCAGCCAT CTGTTGTTTG CCCCTCCCCC
 1210 1220 1230 1240 1250 1260
 GTGCCCTCCT TGACCCCTGGA AGGTGCCACT CCCACTGTCC TTTCCTAATA AAATGAGGAA
 1270 1280 1290 1300 1310 1320
 ATTGCATCGC ATTGTCTGAG TAGGTGTAT TCTATTCTGG GGGGTGGGGT GGGGCAGGAC

FIGURE 7

DNASIS
Desmond

10 / 51

1330 1340 1350 1360 1370 1380
 AGCAAGGGGG AGGATTGGGA AGACAATAGC AGGCATGCTG GGGATGCGGT GGGCTCTATG
 1390 1400 1410 1420 1430 1440
 GCTTCTGAGG CGGAAAGAAC CAGCTGGGC TCGAAGCGGC CGCCCATTT GCTGGTGGTC
 1450 1460 1470 1480 1490 1500
 AGATGCAGGA TGGCGTGGGA CGCGGCGGGG AGCGTCACAC TGAGGTTTC CGCCAGACG
 1510 1520 1530 1540 1550 1560
 CACTGCTGCC AGGCCTGAT GTGCCCGGCT TCTGACCATG CGGTCGCGTT CGGTTGCACT
 1570 1580 1590 1600 1610 1620
 ACGCGTACTG TGAGCCAGAG TTGCCCAGGCT CTCTCCGGCT GCGGTAGTT AGGCAGTTCA
 1630 1640 1650 1660 1670 1680
 ATCAACTGTT TACCTTGTGG AGCGACATCC AGAGGCACCTT CACCGCTTGC CAGCGGCTTA
 1690 1700 1710 1720 1730 1740
 ATCCAGCG CCACCATCCA GTGCAGGAGC TCGTTATCGC TATGACGGAA CAGGTATTG
 1750 1760 1770 1780 1790 1800
 CTGGTCACCTT CGATGGTTTG CCCGGATAAA CGGAACCTGGA AAAACTGCTG CTGGTGTGTT
 1810 1820 1830 1840 1850 1860
 GCTTCCGTCA GCGCTGGATG CGCGTGTGG TCAGCAAAGA CCAGACCGTT CATAACAGAAC
 1870 1880 1890 1900 1910 1920
 TGGCGATCGT TCGGCGTATC GCCAAAATCA CCGCCGTAAG CCGACCCACGG GTTGGCGTTT
 1930 1940 1950 1960 1970 1980
 TCATCATATT TAATCAGCGA CTGATCCACC CAGTCCCAGA CGAAGCCGCC CTGTAAACGG
 1990 2000 2010 2020 2030 2040
 GGATACTGAC GAAACGCCCTG CCAGTATTAA GCGAAACCGC CAAGACTGTT ACCCATCGCG
 2050 2060 2070 2080 2090 2100
 CGCGTATT CGCAAAGGAT CAGCGGGCGC GTCTCTCCAG GTAGCGAAAG CCATTTTTG
 2110 2120 2130 2140 2150 2160
 ATGGACCATT TCGGCACAGC CGGGAAAGGGC TGGTCTTCAT CCACCGCGGC GTACATCGGG
 2170 2180 2190 2200 2210 2220
 CAAATAATAT CGGTGGCCGT GGTGTCGGCT CCGCCGCCCTT CATACTGCAC CGGGCGGGAA
 2230 2240 2250 2260 2270 2280
 GGATCGACAG ATTTGATCCA GCGATAACAGC GCGTCGTGAT TAGCGCCGTG GCCTGATTCA
 2290 2300 2310 2320 2330 2340
 TTCCCCAGCG ACCAGATGAT CACACTCGGG TGATTACGAT CGCGCTGCAC CATTGCGTT
 2350 2360 2370 2380 2390 2400
 ACGCGTTCGC TCATGCCGG TAGCCAGCGC GGATCATCGG TCAGACGATT CATTGGCAC
 2410 2420 2430 2440 2450 2460
 ATGCCGTGGG TTTCAATATT GGCTTCATCC ACCACATACA GGCGTAGCG GTCGCACAGC
 2470 2480 2490 2500 2510 2520
 GTGTACCAACA CGGGATGGTT CGGATAATGC GAACAGCGCA CGGCCTTAAA GTTGGTCTGC
 2530 2540 2550 2560 2570 2580
 TTCATCAGCA GGATATCCTG CACCATCGTC TGCTCATCCA TGACCTGACC ATGCAGAGGA
 2590 2600 2610 2620 2630 2640

DNASIS
Desmond Clark

11 / 51

TGATGCTCGT GACGGTTAAC GCCTCGAATC AGCAACGGCT TGCCTGTTAG CAGCAGCAGA
 2650 2660 2670 2680 2690 2700
 CCATTTCAA TCCGCACCTC GCGGAAACCG ACATCGCAGG CTTCTGCTTC AATCAGCGTG
 2710 2720 2730 2740 2750 2760
 CGCTGGCGG TGTGCAGTTC AACCACCGCA CGATAGAGAT TCGGGATTTC GGCGCTCCAC
 2770 2780 2790 2800 2810 2820
 AGTTTCGGGT TTTCGACGTT CAGACGTAGT GTGACGCGAT CGGCATAACC ACCACGCTCA
 2830 2840 2850 2860 2870 2880
 TCGATAATTT CACCGCCGAA AGGCGCGGTG CCGCTGGCGA CCTGCCTTTC ACCCTGCCAT
 2890 2900 2910 2920 2930 2940
 AAAGAAACTG TTACCCGTAG GTAGTCACGC AACTCGCCGC ACATCTGAAC TTCAGCCTCC
 2950 2960 2970 2980 2990 3000
 AGTACAGCGC GGCTGAAATC ATCATTAAAG CGAGTGGCAA CATGGAAATC GCTGATTGT
 3010 3020 3030 3040 3050 3060
 GAGTCGGTT TATGCAGCAA CGAGACGTCA CGGAAAATGC CGCTCATCCG CCACATATCC
 3070 3080 3090 3100 3110 3120
 TGATCTTCCA GATAACTGCC GTCACTCCAG CGCAGCACCA TCACCGCGAG GCGGTTTTCT
 3130 3140 3150 3160 3170 3180
 CGGGCGCGTA AAAATGCGCT CAGGTCAAAT TCAGACGGCA AACGACTGTC CTGGCCGTAA
 3190 3200 3210 3220 3230 3240
 CCGACCCAGC GCCCGTTGCA CCACAGATGA AACGCCGAGT TAACGCCATC AAAAATAATT
 3250 3260 3270 3280 3290 3300
 CGCGTCTGGC CTTCTGTAG CCAGCTTCA TCAACATTAA ATGTGAGCGA GTAACAACCC
 3310 3320 3330 3340 3350 3360
 GTCGGATTCT CCGTGGGAAC AACGGCGGA TTGACCGTAA TGGGATAGGT CACGTTGGTG
 3370 3380 3390 3400 3410 3420
 TAGATGGCG CATCGTAACC GTGCATCTGC CAGTTGAGG GGACGACGAC AGTATCGGCC
 3430 3440 3450 3460 3470 3480
 TCAGGAAGAT CGCACTCCAG CCAGCTTCC GGCACCGCTT CTGGTGCCGG AAACCAGGCA
 3490 3500 3510 3520 3530 3540
 AACGCCTATT CGCCATTCA GCTGCACAC TGTTGGGAAG GGCGATCGGT GCGGGCCTCT
 3550 3560 3570 3580 3590 3600
 TCGCTATTAC GCCAGCTGGC GAAAGGGGGA TGTGCTGCAA GGCGATTAAG TTGGGTAACG
 3610 3620 3630 3640 3650 3660
 CCAGGGTTTT CCCAGTCACG ACGTTGTAAA ACGACTTAAT CCGTCGAGGG GCTGCCTCGA
 3670 3680 3690 3700 3710 3720
 AGCAAACGAC CTTCCGTTGT GCAGCCAGCG GCGCCTGCGC CGGTGCCAC AATCGTGCAC
 3730 3740 3750 3760 3770 3780
 GAACAAACTA ACCAGAACAA ATTATACCG GCGGCACCGC CGCCACCAAC TTCTCCCCTG
 3790 3800 3810 3820 3830 3840
 CCTAACATTTC CAGCGCTCC ACCACCACCA CCACCATCGA TGTCTGAATT GCGCCCCGCT
 3850 3860 3870 3880 3890 3900
 CCACCAATGC CGACGGAACC TCAACCCGCT GCACCTTAG ACGACAGACAA ACAATTGTTG

DNASIS
Desmond

3910 3920 3930 3940 3950 3960
 GAAGCTATTAA GAAACGAAAA AAATCGCACT CGTCTCAGAC CGGCTCTCTT AAGGTAGCTC
 3970 3980 3990 4000 4010 4020
 AAACCAAAAA CGGCGCCGA ACCAGTACA ATAGTTGAGG TGCCGACTGT GTTGCCTAAA
 4030 4040 4050 4060 4070 4080
 GAGACATTTG AGCCTAAACC GCCGTCTGCA TCACCGCCAC CACCTCCGCC TCCGCCCTCCG
 4090 4100 4110 4120 4130 4140
 CCGCCAGCCC CGCCTGCGCC TCCACCGATG GTAGATTTAT CATCAGCTCC ACCACCGCCG
 4150 4160 4170 4180 4190 4200
 CCATTAGTAG ATTGGCGTC TGAAATGTTA CCACCGCCTG CACCATCGCT TTCTAACGTG
 4210 4220 4230 4240 4250 4260
 TTGTCTGAAT TAAATCGGG CACAGTTAGA TTGAAACCCG CCCAAAAACG CCCGCAATCA
 4270 4280 4290 4300 4310 4320
 ATAATTC CAAAAGCTC AACTACAAAT TTGATCGCGG ACGTGTAGC CGACACAATT
 4330 4340 4350 4360 4370 4380
 AATAGGCGTC GTGTGGCTAT GGCAAAATCG TCTTCGGAAG CAACTTCTAA CGACGAGGGT
 4390 4400 4410 4420 4430 4440
 TGGGACGACG ACGATAATCG GCCTAATAAA GCTAACACGC CCGATGTTAA ATATGTCAA
 4450 4460 4470 4480 4490 4500
 GCTACTAGTG GTACCTTAAT TAAGGGCGG AGAATGGGCG GAACTGGCG GAGTTAGGGG
 4510 4520 4530 4540 4550 4560
 CGGGATGGGC GGAGTTAGGG GCGGGACTAT GGTTGCTGAC TAATTGAGAT GCATGCTTTG
 4570 4580 4590 4600 4610 4620
 CATACTTCTG CCTGCTGGGG AGCCTGGGA CTTTCCACAC CTGGTTGCTG ACTAATTGAG
 4630 4640 4650 4660 4670 4680
 GCATGCTT TGCACTACTTC TGCCGTCTGG GGAGCCTGGG GACTTTCCAC ACCCTAACTG
 4690 4700 4710 4720 4730 4740
 ACACACATTTC CACAGAATTAA ATTCCCCTAG TTATTAATAG TAATCAATTAA CGGGGTCTT
 4750 4760 4770 4780 4790 4800
 AGTTCATAGC CCATATATGG AGTTCCCGGT TACATAACTT ACGGTAAATG GCCCGCCTGG
 4810 4820 4830 4840 4850 4860
 CTGACCGCCCC AACGACCCCCC GCCCATTGAC GTCAATAATG ACGTATGTTCC CCACTAGTAAC
 4870 4880 4890 4900 4910 4920
 GCCAATAGGG ACTTTCCATT GACGTCAATG GGTGGAGTAT TTACGGTAAA CTGCCCCACTT
 4930 4940 4950 4960 4970 4980
 GGCAGTACAT CAAGTGTATC ATATGCCAAG TACGCCCCCT ATTGACGTCA ATGACGGTAA
 4990 5000 5010 5020 5030 5040
 ATGGCCCGCC TGGCATTATG CCCAGTACAT GACCTTATGG GACTTTCTA CTTGGCAGTA
 5050 5060 5070 5080 5090 5100
 CATCTACGTA TTAGTCATCG CTATTACCAT GGTGATGCGG TTTTGGCAGT ACATCAATGG
 5110 5120 5130 5140 5150 5160
 GCGTGGATAG CGGTTTGAET CACGGGGATT TCCAAGTCTC CACCCCATGG ACGTCAATGG
 5170 5180 5190 5200 5210 5220
 GAGTTTGTGTT TGAAGCTTGG CGGGCCAGCT TTATTTAACG TGTTTACGTC GAGTCATTG

DNASIS
Desmond

5230 5240 5250 5260 5270 5280
 TACACTAACG ACAGTGATGA AAGAAATACA AAAGCGATA ATATTTGAA CGACGTGAA
 5290 5300 5310 5320 5330 5340
 CCTTATTAC AAAACAAAAC ACAAACGAAT ATCGACAAAG CTAGATTGCT GCTACAAGAT
 5350 5360 5370 5380 5390 5400
 TTGGCAAGTT TTGTGGCGTT GAGCGAAAAT CCATTAGATA GTCCAGCCAT CGGTTCGGAA
 5410 5420 5430 5440 5450 5460
 AAACAACCCCT TGTTGAAAC TAATCGAAAC CTATTTACA AATCTATTGA GGATTTAATA
 5470 5480 5490 5500 5510 5520
 TTAAATTCA GATATAAAGA CGCTGAAAAT CATTGATTTC TCGCTCTAAC ATACCACCC
 5530 5540 5550 5560 5570 5580
 AAAGATTATA AATTAAATGA ATTATTAAAA TACATCAGCA ACTATATATT GATAGACATT
 5590 5600 5610 5620 5630 5640
 ...AGTTTGT GATATTAGTT TGTGCGTCTC ATTACAATGG CTGTTATTT TAACAACAAA
 5650 5660 5670 5680 5690 5700
 CAACTGCTCG CAGACAATAG TATAGAAAAG GGAGGTGAAC TGTTTTGTT TAACGGTTCG
 5710 5720 5730 5740 5750 5760
 TACAACATTG TGGAAAGTTA TGTTAATCCG GTGCTGCTAA AAAATGGTGT AATTGAACTA
 5770 5780 5790 5800 5810 5820
 GAAGAAGCTG CGTACTATGC CGGCAACATA TTGTACAAAAA CCGACGATCC CAAATTCA
 5830 5840 5850 5860 5870 5880
 GATTATATAA ATTAAATAAT TAAAGCAACA CACTCCGAAG AACTACCAGA AAATAGCACT
 5890 5900 5910 5920 5930 5940
 GTTGTAAATT ACAGAAAAAC TATGCGCAGC GGTACTATAC ACCCCATTAA AAAAGACATA
 5950 5960 5970 5980 5990 6000
 ...TATTATG ACAACAAAAA ATTACTCTA TACGATAGAT ACATATATGG ATACGATAAT
 6010 6020 6030 6040 6050 6060
 AACTATGTTA ATTITATGA GGAGAAAAAT GAAAAGAGA AGGAATACGA AGAAGAAGAC
 6070 6080 6090 6100 6110 6120
 GACAAGGCCTG CTAGTTTATG TGAAAATAAA ATTATATTGT CGCAAATTAA CTGTGAATCA
 6130 6140 6150 6160 6170 6180
 TTTGAAAATG ATTAAATAA TTACCTCAGC GATTATAACT ACGCGTTTC AATTATAGAT
 6190 6200 6210 6220 6230 6240
 AATACTACAA ATGTTCTTGT TGCGTTGGT TTGTATCGTT AATAAAAAC AAATTGACA
 6250 6260 6270 6280 6290 6300
 TTTATAATTG TTTTATTATT CAATAATTAC AAATAGGATT GAGACCCCTG CAGTTGCCAG
 6310 6320 6330 6340 6350 6360
 CAAACGGACA GAGCTTGTG AGGAGAGTTG TTGATTCACT GTTGCCTCC CTGCTGCGGT
 6370 6380 6390 6400 6410 6420
 TTTTCACCGA AGTCATGCC AGTCCAGCGT TTTTGCAGCA GAAAAGCCGC CGACTTCGGT
 6430 6440 6450 6460 6470 6480
 TTGCGGTCGC GAGTGAAGAT CCCTTTCTTG TTACCGCCAA CGCGCAATAT GCCTTGCAG
 6490 6500 6510 6520 6530 6540

DNASIS
Desmond

GTCGCAAAAT CGGCATAATT CCATACCTGT TCACCGACGA CGGCCTGAC GCGATCAAAG

6550 6560 6570 6580 6590 6600
ACGGGTGAT ACATATCCAG CCATGCACAC TGATACTCTT CACTCCACAT GTCGGTGAC

6610 6620 6630 6640 6650 6660
ATTGAGTGCA GCCCGGCTAA CGTATCCACG CCGTATTGG TGATGATAAT CGGCTGATGC

6670 6680 6690 6700 6710 6720
AGTTTCTCCT GCCAGGCCAG AAGTTCTTT TCCAGTACCT TCTCTGCCGT TTCCAAATCG

6730 6740 6750 6760 6770 6780
CCGCTTTGGA CATAACCACCC GTAATAACGG TTCAGGCACA GCACATCAA GAGATCGCTG

6790 6800 6810 6820 6830 6840
ATGGTATCGG TGTGAGCGTC GCAGAACATT ACATTGACGC AGGTGATCGG ACGCGTCGGG

6850 6860 6870 6880 6890 6900
TCGAGTTTAC GCGTTGCTTC CGCCAGTGGC GCGAAATATT CCCGTGCACC TTGCGGACGG

6910 6920 6930 6940 6950 6960
GTATCCGGTT CGTTGGCAAT ACTCCACATC ACCACGCTT GGTGGTTTTT GTCACGCGCT

6970 6980 6990 7000 7010 7020
ATCAGCTCTT TAATCGCCTG TAAGTGCCTG TGCTGAGTTT CCCCCTTGAC TGCCCTTTCG

7030 7040 7050 7060 7070 7080
CTGTACAGTT CTTTCGGCTT GTTGCCCCGT TCGAAACCAA TGCCCTAAAGA GAGGTTAAAG

7090 7100 7110 7120 7130 7140
CCGACAGCAG CAGTTTCATC AATCACCAACG ATGCCATGTT CATCTGCCA GTCGAGCATC

7150 7160 7170 7180 7190 7200
TCTTCAGCGT AAGGGTAATG CGAGGTACGG TAGGAGTTGG CCCCCATCCA GTCCATTAAT

7210 7220 7230 7240 7250 7260
GCGTGGTCGT GCACCATCAG CACGTTATCG AATCCTTTGC CACGCAAGTC CGCATCTTC

7270 7280 7290 7300 7310 7320
TGACGACCAA AGCCAGTAAA GTAGAACGGT TTGTGGTTAA TCAGGAACGT TTCGCCCTTC

7330 7340 7350 7360 7370 7380
ACTGCCACTG ACCGGATGCC GACGCGAAGC GGGTAGATAT CACACTCTGT CTGGCTTTG

7390 7400 7410 7420 7430 7440
GCTGTGACGC ACAGTTCATC GAGATAACCT TCACCCGGTT GCCAGAGGTG CGGATTCAACC

7450 7460 7470 7480 7490 7500
ACTTGCAAAG TCCCGCTAGT GCCTTGTCCA GTTGCAACCA CCTGTTGATC CGCATCACGC

7510 7520 7530 7540 7550 7560
AGTTCAACGC TGACATCACC ATTGGCCACC ACCTGCCAGT CAACAGACGC GTGGTTACAG

7570 7580 7590 7600 7610 7620
TCTTGCGCGA CATGCGTCAC CACGGTGATA TCGTCCACCC AGGTGTTGG CGTGGTGAG

7630 7640 7650 7660 7670 7680
AGCATTACGC TGCAGATGGAT TCCGGCATAG TAAAGAAAT CATGGAAGTA AGACTGCTTT

7690 7700 7710 7720 7730 7740
TTCTTGCCGT TTTCGTCGGT AATCACCATT CCCGGCGGGA TAGTCTGCCA GTTCAGTTCG

7750 7760 7770 7780 7790 7800
TTGTTCACAC AAACGGTGAT ACCCCTCGAC GGATTAAGA CTTCAAGCGG TCAACTATGA

DNASIS
Desmond

7810 7820 7830 7840 7850 7860
 AGAAAGTGTTC GTCTTCGTCC CAGTAAGCTA TGTCTCCAGA ATGTAGGCCAT CCATCCTTGT
 7870 7880 7890 7900 7910 7920
 CAATCAAGGC GTTGGTCGCT TCCGGATTGT TTACATAACC GGACATAATC ATAGGTCCCTC
 7930 7940 7950 7960 7970 7980
 TGACACATAA TTCGCCTCTC TGATTAACGC CCAGCGTTTT CCCGGTATCC AGATCCACAA
 7990 8000 8010 8020 8030 8040
 CCTTCGCTTC AAAAAATGGA ACAACTTAC CGACCGCGCC CGGTTTATCA TCCCCCTCGG
 8050 8060 8070 8080 8090 8100
 GTGTAATCAG AATAGCTGAT GTAGTCTCAG TGAGCCCATA TCCTTGTGTT ATCCCTGGAA
 8110 8120 8130 8140 8150 8160
 GATGGAAGCG TTTTGCACCC GCTTCCCCGA CTTCTTCGA AAGAGGTGCG CCCCCAGAAG
 8170 8180 8190 8200 8210 8220
 1TTTCGTG TAAATTAGAT AAATCGTATT TGTCAATCAG AGTGTCTTGT GCGAAGAACG
 8230 8240 8250 8260 8270 8280
 AAAATAGGGT TGGTACTAGC AACGCACTTT GAATTTTGTA ATCCTGAAGG GATCGTAAAA
 8290 8300 8310 8320 8330 8340
 ACAGCTCTTC TTCAAATCTA TACATTAAGA CGACTCGAAA TCCACATATC AAATATCCGA
 8350 8360 8370 8380 8390 8400
 GTGTAGTAAA CATTCCAAAA CCGTGATGGA ATGGAACAAC ACTTAAATC GCAGTATCCG
 8410 8420 8430 8440 8450 8460
 GAATGATTTG ATTGCCAAAA ATAGGATCTC TGGCATGCGA GAATCTGACG CAGGCAGTTC
 8470 8480 8490 8500 8510 8520
 TATGCGGAAG GGCCACACCC TTAGGTAAAC CAGTAGATCC AGAGGAATTG TTTTGTACG
 8530 8540 8550 8560 8570 8580
 CAAAGGAC TCTGGTACAA AATCGTATTTC ATTAAAACCG GGAGGTAGAT GAGATGTGAC
 8590 8600 8610 8620 8630 8640
 GAACGTGTAC ATCGACTGAA ATCCCTGGTA ATCCGTTTTA GAATCCATGA TAATAATT
 8650 8660 8670 8680 8690 8700
 CTGGATTATT GGTAATTTTT TTTGCACGTT CAAAATTTTG TGCAACCCCT TTTTGGAAAC
 8710 8720 8730 8740 8750 8760
 AACACACTACG GTAGGCTGCG AAATGTTCAT ACTGTTGAGC AATTCACTT CATTATAAT
 8770 8780 8790 8800 8810 8820
 GTCGTTCGCG GGGCGCACTG CAACTCCGAT AAATAACGCG CCCAACACCG GCATAAAGAA
 8830 8840 8850 8860 8870 8880
 TTGAAGAGAG TTTTCACTGC ATACGACGAT TCTGTGATTG GTATTCAAGCC CATATCGTT
 8890 8900 8910 8920 8930 8940
 CATAGCTTCT GCCAACCGAA CGGACATTTC GAAGTATTCC GCGTACGTGA TGTTCACCTC
 8950 8960 8970 8980 8990 9000
 GATATGTGCA TCTGTAAAAG GAATTGTTCC AGGAACCAGG GCGTATCTCT TCATAGCCTT
 9010 9020 9030 9040 9050 9060
 ATGCAGTTGC TCTCCAGCGG TTCCATCCTC TAGCTTGTCT TCTCAATTTC TTATTGCA
 9070 9080 9090 9100 9110 9120
 AATGAGAAAA AAAGGAAAAT TAATTTAAC ACCAATTCAAG TAGTTGATTG AGCAAATGCG

DNASIS
Desmond

16 / 51

9130 9140 9150 9160 9170 9180
 TTGCCAAAAA GGATGCTTA GAGACAGTGT TCTCTGCACA GATAAGGACA AACATTATTC
 9190 9200 9210 9220 9230 9240
 AGAGGGAGTA CCCAGAGCTG AGACTCCTAA GCCAGTGAGT GGCACAGCAT CCAGGGAGAA
 9250 9260 9270 9280 9290 9300
 ATATGCTTGT CATCACCGAA GCCTGATTCC GTAGAGCCAC ACCCTGGTAA GGGCCAATCT
 9310 9320 9330 9340 9350 9360
 GCTCACACAG GATAGAGAGG GCAGGAGCCA GGGCAGAGCA TATAAGGTGA GGTAGGATCA
 9370 9380 9390 9400 9410 9420
 GTTGCTCCTC ACATTTGCTT CTGACATAGT TGTGTTGGGA GCTTGGATCG ATCCACCATG
 9430 9440 9450 9460 9470 9480
 GGCTTCAATA CCCTGATTGA CTGGAACAGC TGTAGCCCTG AACAGCAGCG TGCGCTGCTG
 9490 9500 9510 9520 9530 9540
 ACGTCCGG CGATTTCCGC CTCTGACAGT ATTACCCGGA CGGTAGCGA TATTTGGAT
 9550 9560 9570 9580 9590 9600
 AATGTAAAAA CGCGCGGTGA CGATGCCCTG CGTGAATACA GCGCTAAATT TGATAAAACA
 9610 9620 9630 9640 9650 9660
 GAAGTGACAG CGCTACGCGT CACCCCTGAA GAGATGCCCG CGGCCGGCGC GCGTCTGAGC
 9670 9680 9690 9700 9710 9720
 GACGAATTAA AACAGGGCAT GACCGCTGCC GTCAAAAATA TTGAAACGTT CCATTCGCG
 9730 9740 9750 9760 9770 9780
 CAGACGCTAC CGCCTGTAGA TGTGGAAACC CAGCCAGGCG TGCGTTGCCA GCAGGTTACG
 9790 9800 9810 9820 9830 9840
 CGTCCCCTCT CGTCTGTCGG TCTGTATATT CCCGGCGGCT CGGCTCCGCT CTTCTAACG
 9850 9860 9870 9880 9890 9900
 CTGATGC TGGCGACGCC GGCGCGCATT GCGGGATGCC AGAAGGTGGT TCTGTGCTG
 9910 9920 9930 9940 9950 9960
 CCGCCGCCCA TCGCTGATGA AATCCTCTAT GCGGGCGAAC TGTGTGGCGT GCAGGAAATC
 9970 9980 9990 10000 10010 10020
 TTTAACGTGCG GCGGCGCGCA GGCGATTGCC GCTCTGGCCT TCGGCAGCGA GTCCGTACCG
 10030 10040 10050 10060 10070 10080
 AAAGTGGATA AAATTTTGG CCCCCGGCAAC GCCTTTGTAA CGAAGCCAA ACGTCAAGGTC
 10090 10100 10110 10120 10130 10140
 AGCCAGCGTC TCGACGGCGC GGCTATCGAT ATGCCAGCCG GGCGTCTGA AGTACTGGTG
 10150 10160 10170 10180 10190 10200
 ATCGCAGACA GCGGGCGAAC ACCGGATTC GTCGCTCTG ACCTGCTCTC CCAGGCTGAG
 10210 10220 10230 10240 10250 10260
 CACGGCCCGG ATTCCCAGGT GATCCTGCTG ACGCCTGATG CTGACATTGC CCGCAAGGTG
 10270 10280 10290 10300 10310 10320
 GCGGAGGCAG TAGAACGTCA ACTGGCGGAA CTGCCCGCG CGGACACCGC CCGGCAGGCC
 10330 10340 10350 10360 10370 10380
 CTGAGCGCCA GTCGTCTGAT TGTGACCAAA GATTTAGCGC AGTGCCTCGC CATCTCTAAT
 10390 10400 10410 10420 10430 10440

DNASIS
Desmond Park

CAGTATGGGC CGGAACACTT AATCATCCAG ACGCGCAATG CGCGCGATTG GGTGGATGCG
 10450 10460 10470 10480 10490 10500
 ATTACCAGCG CAGGCTCGGT ATTTCTCGGC GACTGGTCGC CGGAATCCGC CGGTGATTAC
 10510 10520 10530 10540 10550 10560
 GCTTCCGAA CCAACCATGT TTTACCGACC TATGGCTATA CTGCTACCTG TTCCAGCCTT
 10570 10580 10590 10600 10610 10620
 GGGTTAGCGG ATTTCCAGAA ACGGATGACC GTTCAGGAAC TGTGAAAGC GGGCTTTTCC
 10630 10640 10650 10660 10670 10680
 GCTCTGGCAT CAACCATTGA AACATTGGCG CGGGCAGAAC GTCTGACCGC CCATAAAAAT
 10690 10700 10710 10720 10730 10740
 GCCGTGACCC TGCGCGTAAA CGCCCTCAAG GAGCAAGCAT GAGCACTGAA AACACTCTCA
 10750 10760 10770 10780 10790 10800
 CGCTCGCTGA CTTAGCCCCT GAAAATGTCC GCAACCTGGA GATCCAGACA TGGATAAGAT
 10810 10820 10830 10840 10850 10860
 ACATTGATGA GTTTGGACAA ACCACAACTA GAATGCAGTG AAAAAAAATGC TTTATTTGTG
 10870 10880 10890 10900 10910 10920
 AAATTTGTGA TGCTATTGCT TTATTTGTAA CCATTATAAG CTGCAATAAA CAAGTTAAC
 10930 10940 10950 10960 10970 10980
 ACAACAATTG CATTCACTTT ATGTTTCAGG TTCAGGGGGG GGTGTGGGAG GTTTTTTAAA
 10990 11000 11010 11020 11030 11040
 GCAAGTAAAA CCTCTACAAA TGTGGTATGG CTGATTATGA TCTCTAGGGC CGGCCCTCGA
 11050 11060 11070 11080 11090 11100
 CGGCGCGCCT GGCGCTACT AACTCTCTCC TCCCTCCTT TTCCCTGCAGG CTCAAGGC
 11110 11120 11130 11140 11150 11160
 GCATGCCCGA CGGGGAGGAT CTCGTCGTGA CCCATGGCGA TGCCCTGCTTG CCGAATATCA
 11170 11180 11190 11200 11210 11220
 TGGTGGAAAA TGGCCGCTT TCTGGATTCA TCGACTGTGG CCGGCTGGGT GTGGCGGACC
 11230 11240 11250 11260 11270 11280
 GCTATCAGGA CATAGCGTTG GCTACCCGTG ATATTGCTGA AGAGCTTGGC GGCAGATGGG
 11290 11300 11310 11320 11330 11340
 CTGACCGCTT CCTCGTGCTT TACGGTATCG CCCCTCCGA TTCGCAGCGC ATCGCTTCT
 11350 11360 11370 11380 11390 11400
 ATCGCCTTCT TGACGGAGTTC TTCTGAGCGG GACTCTGGGG TTCGAAATGA CCGACCAAGC
 11410 11420 11430 11440 11450 11460
 GACGCCAAC CTGCCATCAC GAGATTCGA TTCCACCGCC GCCTTCTATG AAAGGTTGGG
 11470 11480 11490 11500 11510 11520
 CTTCGGAATC GTTTTCCGGG ACGCCGGCTG GATGATCCTC CAGCGCGGGG ATCTCATGCT
 11530 11540 11550 11560 11570 11580
 GGAGTTCTTC GCCCACCCA ACTTGTTAT TGCAAGCTTAT AATGGTTACA AATAAGCAA
 11590 11600 11610 11620 11630 11640
 TAGCATCACA AATTCACAA ATAAGCATT TTTTCACTG CATTCTAGTT GTGGTTTGT
 11650 11660 11670 11680 11690 11700
 CAAACTCATC AATCTATCTT ATCATGTCTG GATCGCGGCC GGTCTCTCTC TAGCCCTAGG

DNASIS
Desmond Lark

18 / 51

11710 11720 11730 11740 11750 11760
 TCTAGACTTG GCAGAACATA TCCATCGCGT CCGCCATCTC CAGCAGCCGC ACGCGGCCGA

 11770 11780 11790 11800 11810 11820
 TCTCGGGCAG CGTTGGGTCC TGGCCACGGG TGCGCATGAT CGTGCCTCTG TCGTTGAGGA

 11830 11840 11850 11860 11870 11880
 CCCGGCTAGG CTGGCGGGGT TGCCCTACTG GTTAGCAGAA TGAATCACCG ATACCGGAGC

 11890 11900 11910 11920 11930 11940
 GAACGTGAAG CGACTGCTGC TGCAAAACGT CTGCACCTG AGCAACAAACA TGAATGGTCT

 11950 11960 11970 11980 11990 12000
 TCGGTTTCGG TGTTTCGTA AGTCTGGAAA CGCGGAAGTC AGCGCCCTGC ACCATTATGT

 12010 12020 12030 12040 12050 12060
 TCCGGATCTG CATCGCAGGA TGCTGCTGGC TACCCCTGTGG AACACCTACA TCTGTATTAA

 12070 12080 12090 12100 12110 12120
 CGAACCGCTG GCATTGACCC TGAGTGATTT TTCTCTGGTC CCGCCGCATC CATAACGCCA

 12130 12140 12150 12160 12170 12180
 GTTGTITACC CTCACAAACGT TCCAGTAACC GGGCATGTTC ATCATCAGTA ACCCGTATCG

 12190 12200 12210 12220 12230 12240
 TGAGCATCCT CTCTCGTTTC ATCGGTATCA TTACCCCCAT GAACAGAAAT CCCCCCTTACA

 12250 12260 12270 12280 12290 12300
 CGGAGGCATC AGTGACCAAA CAGGAAAAAA CCGCCCTTAA CATGGCCCAGC TTTATCAGAA

 12310 12320 12330 12340 12350 12360
 GCCAGACATT AACGCTTCTG GAGAAACTCA ACGAGCTGGA CGCGGATGAA CAGGCAGACA

 12370 12380 12390 12400 12410 12420
 TCTGTGAATC GCTTCACGAC CACCGCTGATG AGCTTTACCG CAGCTGCCCTC GCGCGTTTCG

 12430 12440 12450 12460 12470 12480
 GTGATGACGG TGAAAACCTC TGACACATGC AGCTCCCGA GACGGTCACA GCTTGTCTGT

 12490 12500 12510 12520 12530 12540
 AAGCGGATGC CGGGAGCAGA CAAGCCCCGTC AGGGCGCGTC AGCGGGTGT GGCGGGTGTG

 12550 12560 12570 12580 12590 12600
 GGGGCGCAGC CATGACCCAG TCACGTAGCG ATAGCGGAGT GTATACTGGC TTAACTATGC

 12610 12620 12630 12640 12650 12660
 GGCATCAGAG CAGATTGTAC TGAGAGTCCA CCATATGCCG TGTGAAATAC CGCACAGATG

 12670 12680 12690 12700 12710 12720
 CGTAAGGAGA AAATACCGCA TCAGGCGCTC TTCCGCTTCC TCGCTCACTG ACTCGCTGCG

 12730 12740 12750 12760 12770 12780
 CTCGGTCGTT CGGCTGCCGC GAGCGGTATC AGCTCACTCA AAGGCGGTAA TACGGTTATC

 12790 12800 12810 12820 12830 12840
 CACAGAACATCA GGGGATAAACG CAGGAAAGAA CATGTGAGCA AAAGGCCAGC AAAAGGCCAG

 12850 12860 12870 12880 12890 12900
 GAACCGTAAA AAGGCCGCGT TGCTGGCGTT TTCCCATAGG CTCCGCCCCC CTGACGAGCA

 12910 12920 12930 12940 12950 12960
 TCACAAAAAT CGACGCTCAA GTCAGAGGTG GCGAAACCCG ACAGGACTAT AAAGATAACCA

 12970 12980 12990 13000 13010 13020
 GGC GTTTCCC CCTGGAAGCT CCCTCGTGC CGTCTCTGTT CCGACCCCTGC CGCTTACCGG

DNASIS
Desmond

13030	13040	13050	13060	13070	13080
ATACCTGTCC GCCTTCTCC CTTGGGAAG CGTGGCGTT TCTCATAGCT CACGCTGTAG					
13090	13100	13110	13120	13130	13140
GTATCTCA GT TCGGTGTAGG TCGTTCGCTC CAAGCTGGC TGTTGCACG AACCCCCCGT					
13150	13160	13170	13180	13190	13200
TCAGCCCAC CGCTGCGCCT TATCCGGTAA CTATCGTCTT GAGTCCAACC CGGTAAGACA					
13210	13220	13230	13240	13250	13260
CGACTTATCG CCACTGGCAG CAGCCACTGG TAACAGGATT AGCAGAGCGA GGTATGTAGG					
13270	13280	13290	13300	13310	13320
CGGTGCTACA GAGTTCTTGA AGTGGTGCC TAAC TACGGC TACACTAGAA GGACAGTATT					
13330	13340	13350	13360	13370	13380
TGGTATCTGC GCTCTGCTGA AGCCAGTTAC CTTCGAAAAA AGAGTTGGTA GCTCTTGATC					
13390	13400	13410	13420	13430	13440
. CAAACAA ACCACCGCTG GTAGCGGTGG TTTTTTGTT TGCAAGCAGC AGATTACGCG					
13450	13460	13470	13480	13490	13500
CAGAAAAAAA GGATCTCAAG AAGATCCTTT GATCTTTCT ACGGGGTCTG ACGCTCAGTG					
13510	13520	13530	13540	13550	13560
GAACGAAAAC TCACGTTAAG GGATTTGGT CATGAGATTAA TCAAAAGGA TCTTCACCTA					
13570	13580	13590	13600	13610	13620
GATCCTTTTA AATTAAAAAT GAAGTTTAA ATCAATCTAA AGTATATATG AGTAAACTTG					
13630	13640	13650	13660	13670	13680
GTCTGACAGT TACCAATGCT TAATCGTGA GGCACCTATC TCAGCGATCT GTCTATTTCG					
13690	13700	13710	13720	13730	13740
TTCATCCATA GTTGCTGAC TCCCCGTGCT GTAGATAACT ACGATAACGGG AGGGCTTACC					
13750	13760	13770	13780	13790	13800
. CTGGCCCC AGTGCTGCAA TGATACCGCG AGACCCACGC TCACCGGCTC CAGATTATC					
13810	13820	13830	13840	13850	13860
AGCAATAAAC CAGCCAGCCG GAAGGGCCGA GCGCAGAAGT GGTCTGCAA CTTTATCCGC					
13870	13880	13890	13900	13910	13920
CTCCATCCAG TCTATTAATT GTTGCCGGGA AGCTAGAGTA AGTAGTTCGC CAGTTAATAG					
13930	13940	13950	13960	13970	13980
TTTGCACAC GTTGTGCCA TTGCTGCAGG CATCGTGGTG TCACGCTCGT CGTTTGGTAT					
13990	14000	14010	14020	14030	14040
GGCTTCATTC AGCTCCGGTT CCCAACGATC AAGGCAGATT ACATGATCCC CCATGTTGTG					
14050	14060	14070	14080	14090	14100
CAAAAGCG GTTAGCTCCT TCGGTCTCC GATCGTTGTC AGAAGTAAGT TGGCCGAGT					
14110	14120	14130	14140	14150	14160
GTTATCACTC ATGGTTATGG CAGCACTGCA TAATTCTCTT ACTGTCTGC CATCCGTAAG					
14170	14180	14190	14200	14210	14220
ATGCTTTCT GTGACTGGTG AGTACTCAAC CAAGTCATTC TGAGAATAGT GTATGCGCG					
14230	14240	14250	14260	14270	14280
ACCGAGTTGC TCTTGGCCCG CGTCAACACG GGATAATACC GCGCACATA GCAGAACTTT					
14290	14300	14310	14320	14330	14340

DNASIS
Desmond rk

AAAAGTGCTC ATCATTGGAA AACGTTCTTC GGGGCAGAAA CTCTCAAGGA TCTTACCGCT

14350 14360 14370 14380 14390 14400
GTTGAGATCC AGTTGATGT AACCCACTCG TGCACCCAAC TGATCTTCAG CATCTTTAC

14410 14420 14430 14440 14450 14460
TTTCACCAGC GTTTCTGGGT GAGCAAAAC AGGAAGGCAA AATGCCGCAA AAAAGGGAAT

14470 14480 14490 14500 14510 14520
AAGGGCGACA CGGAAATGTT GAATACTCAT ACTCTTCCTT TTTCAATATT ATTGAAGCAT

14530 14540 14550 14560 14570 14580
TTATCAGGGT TATTGTCCTA TGAGCGGATA CATATTGAA TGTATTTAGA AAAATAAACCA

14590 14600 14610 14620 14630 14640
AATAGGGGTT CCGCGCACAT TTCCCCGAAA AGTGCCACCT GACGTCTAAG AAACCATTAT

14650 14660 14670 14680 14690 14700
TATCATGACA TTAACCTATA AAAATAGGCG TATCACGAGG CCCTTTGTC TTCAGAA..

DNASIS
Molly

FIGURE 8

2/3/98

10 20 30 40 50 60
 TTAATTAAGG GGC GGAGAAT GGG CGGA ACT GGG CGG AGT T A GGG CGG AGT
 70 80 90 100 110 120
 TAG GGG CGGG ACT ATGG TTG CTG ACT ATT GAG ATG C ATG CTT TG C ATAC TT CTG CCT GC
 130 140 150 160 170 180
 TGG GAG GCCT GGG GACT TTTC CAC ACCT GGT TG CTG ACT AA TTG AGAT GCA TG CTT TG CAT
 190 200 210 220 230 240
 ACT TT CTG C CT GCT GGG GAGC CT GGG GACT TCC ACAC CCCT AACT GAC ACA CATT CCAC AG
 250 260 270 280 290 300
 AATTA ATTCC CCTAG TTATT AATAG TAATC AATTAC GGGG TCATTAG TTTC ATAG CCC ATA
 310 320 330 340 350 360
 TATGG AGTT C CGCG TTAC AT AAC TTAC GGT AA ATGG CCCG CCT GGCT GAC CG CCC AAC GA
 370 380 390 400 410 420
 :CCG CCCA TTGAC GTCAA TAATGAC GTA TG TTCC CATA GTA AC GCAA TAGGG ACT TT
 430 440 450 460 470 480
 CCATTGAC GT CAATGG GTGG AGT ATT TAAG GTAA ACTG GCC CACTT GGCA GAG TAC ATCA AGT
 490 500 510 520 530 540
 GTATCATATG CCAAGTAC GC CCC TATT GA CGT CAAT GAC GGT AA ATGG GC CC GCGT GGCA
 550 560 570 580 590 600
 TTATGCC CAG TAC ATGAC CT TATGG GACT T CC TACT TTGG CAG TAC ATCT ACG TATT AGT
 610 620 630 640 650 660
 CAT CGCT ATT ACC ATGG TGA TG CGG TTTTG GCAG TAC ATC AATGG CGT G GATAG CGG TT
 670 680 690 700 710 720
 TGACTCAC CGG GGAT TTCC AA GTCT CCAC CC CATTGAC GTC AATGG AGTT TG TTT GAG
 730 740 750 760 770 780
 TGG CCGG C CAG CT TT ATT TAAC GTG TTT ACG TCG AGTC AATT GTAC AC TAAC GAC AGT
 790 800 810 820 830 840
 GATGAA AGAA ATAC AAA AGC GCATA AT ATT TTGAA CGAC G TGAA CC TT ATTAC AAA AC
 850 860 870 880 890 900
 AAAACACAAA CGA ATAT CGA CAA AGCT AGA TTG CTG CTAC AAG AT TTGG C AAG TTT GTG
 910 920 930 940 950 960
 GCG TTG AGCG AAA ATCC ATT AGA TAGT CCA GCC ATCG GTT CGG AAAA ACA ACC TTG TT
 970 980 990 1000 1010 1020
 GAA ACTA ATC GAA ACCT ATT TTACA AAT CT ATT GAG GATT TAAT ATT TAA ATT CAG AT
 1030 1040 1050 1060 1070 1080
 AAAGAC GCTG AAA ATC ATT GAT TTT CGCT CTA ACAT ACC ACC TAA AGA TTATA AAT TT
 1090 1100 1110 1120 1130 1140
 AATG AATT AT TAA AATAC AT CAG CA ACT AT ATT GAT AG AC AT TTCC AG TTT GTG AT
 1150 1160 1170 1180 1190 1200
 TAG TTT GTGC GTCT CATTAC AATGG CTG TT ATT TTAA ACA ACA AAC AACT GCT CGC AGAC
 1210 1220 1230 1240 1250 1260
 AATAGT ATAG AAA AGG GAGG TGA CTG TTG TTAA CG GTC GTAC AA CATT TTG GAA
 1270 1280 1290 1300 1310 1320
 AGT TAT GTTA ATCC CGG TGCT GCT AAAA AT GGT GTAA TTG AACT AGA AAGA AGC TGC GTAC

DNASIS
Molly

1330 1340 1350 1360 1370 1380
 TATGCCGGCA ACATATTGTA CAAAACCGAC GATCCCAAAT TCATTGATTA TATAAATTAA
 1390 1400 1410 1420 1430 1440
 ATAATTAAAG CAACACACTC CGAAGAACTA CCAGAAAATA GCACTGTGT AAATTACAGA
 1450 1460 1470 1480 1490 1500
 AAAACTATGC GCAGCGGTAC TATACACCCCC ATTAAAAAAG ACATATATAT TTATGACAAC
 1510 1520 1530 1540 1550 1560
 AAAAAATTAA CTCTATACGA TAGATACATA TATGGATACG ATAATAACTA TGTTAATTAA
 1570 1580 1590 1600 1610 1620
 TATGAGGAGA AAAATGAAAA AGAGAAGGAA TACGAAGAAG AAGACGACAA GGC GTCTAGT
 1630 1640 1650 1660 1670 1680
 TTATGTGAAA ATAAAATTAT ATTGTGCAA ATTAACGTG AATCATTTGA AAATGATTAA
 1690 1700 1710 1720 1730 1740
 AAATATTACC TCAGCGATTA TAACTACGCG TTTCAATTAA TAGATAATAC TACAAATGTT
 1750 1760 1770 1780 1790 1800
 CTTGTTGCGT TTGGTTTGTA TCGTTAATAA AAAACAAATT TGACATTAT AATTGTTTAA
 1810 1820 1830 1840 1850 1860
 TTATTCAATA ATTACAAATA GGATTGAGAC CCTTGCAGTT GCCAGCAAAC GGACAGAGCT
 1870 1880 1890 1900 1910 1920
 TGTGAGGAG AGTTGTTGAT TCATTGTTG CCTCCCTGCT GCGGTTTTTC ACCGAAGTTC
 1930 1940 1950 1960 1970 1980
 ATGCCAGTCC AGCGTTTTG CAGCAGAAAA GCCGCCGACT TCGGTTGCG GTCCCGAGTG
 1990 2000 2010 2020 2030 2040
 AAGATCCCTT TCTTGTACCC GCCAACGCGC AATATGCCTT GCGAGGTCGC AAAATCGGCG
 2050 2060 2070 2080 2090 2100
 AAATTCATA CCTGTTCAAC GACGACGGCG CTGACGCGAT CAAAGACGCG GTGATAACATA
 2110 2120 2130 2140 2150 2160
 TCCAGCCATG CACACTGATA CTCTTCACTC CACATGTCGG TGTACATTGA GTGCAGCCCG
 2170 2180 2190 2200 2210 2220
 GCTAACGTAT CCACGCCGTA TTCGGTGATG ATAATCGGCT GATGCAGTTT CTCCCTGCCAG
 2230 2240 2250 2260 2270 2280
 GCCAGAAGTT CTTTTCCAG TACCTTCTCT GCCGTTCCA AATCGCCGCT TTGGACATAC
 2290 2300 2310 2320 2330 2340
 CATCCGTAAT AACGGTTTAG GCACAGCACA TCAAAGAGAT CGCTGATGGT ATCGGTGTGA
 2350 2360 2370 2380 2390 2400
 GCGTCGCGAGA ACATTACATT GACGCAGGTG ATCGGACGCG TCGGGTCGAG TTTACGCGTT
 2410 2420 2430 2440 2450 2460
 GCTTCCGCCA GTGGCGCGAA ATATTCCCGT GCACCTTGC GACGGGTATC CGGTTCGTTG
 2470 2480 2490 2500 2510 2520
 GCAATACTCC ACATCACCAC GCTTGGGTGG TTTTGTAC GCGCTATCAG CTCTTTAATC
 2530 2540 2550 2560 2570 2580
 GCCTGTAAGT GCGCTTGCTG AGTTTCCCCG TTGACTGCCT CTTCGCTGTA CAGTTCTTC
 2590 2600 2610 2620 2630 2640

DNASIS
Molly L

GGCTTGTGCA CCGCTTCGAA ACCAATGCCT AAAGAGAGGT TAAAGCCGAC AGCAGCAGTT

2650 2660 2670 2680 2690 2700
TCATCAATCA CCACGATGCC ATGTTCATCT GCCCAGTCGA GCATCTCTTC AGCGTAAGGG

2710 2720 2730 2740 2750 2760
TAATGCGAGG TACGGTAGGA GTTGGCCCCA ATCCAGTCCA TTAATGCGTG GTCGTGCACC

2770 2780 2790 2800 2810 2820
ATCAGCACGT TATCGAACCT TTTGCCACGC AAGTCCGCAT CTTCATGACG ACCAAAGCCA

2830 2840 2850 2860 2870 2880
GTAAAGTAGA ACGGTTGTG GTTAATCAGG AACTGTTGC CTTTCACTGC CACTGACCGG

2890 2900 2910 2920 2930 2940
ATGCCGACGC GAAGCGGGTA GATATCACAC TCTGTCTGGC TTTGGCTGT GACGCACAGT

2950 2960 2970 2980 2990 3000
-ATAGAGAT AACCTTCACC CGGTTGCCAG AGGTGCGGAT TCACCACTTG CAAAGTCCCG

3010 3020 3030 3040 3050 3060
CTAGTGCCTT GTCCAGTTGC AACCACCTGT TGATCCGCAT CACGCAGTTC AACGCTGACA

3070 3080 3090 3100 3110 3120
TCACCATTGG CCACCACCTG CCAGTCAACA GACGGTGGT TACAGTCTTG CGCGACATGC

3130 3140 3150 3160 3170 3180
GTCACCACGG TGATATCGTC CACCCAGGTG TTCGGCGTGG TGTAGAGCAT TACGCTGCAG

3190 3200 3210 3220 3230 3240
TGGATTCCGG CATAGTTAAA GAAATCATGG AAGTAAGACT GCTTTTCTT GCCGTTTTCG

3250 3260 3270 3280 3290 3300
TCGGTAATCA CCATTCGG CGGGATAGTC TGCCAGTTCA GTTCGTTGTT CACACAAACG

3310 3320 3330 3340 3350 3360
-TGTACCCCC TCGACGGATT AAAGACTTCA AGCGGTCAAC TATGAAGAAG TGTCGTCTT

3370 3380 3390 3400 3410 3420
CGTCCCAGTA AGCTATGTCT CCAGAATGTA GCCATCCATC CTTGTCAATC AAGGC GTTGG

3430 3440 3450 3460 3470 3480
TCGCTTCCGG ATTGTTTACA TAACCGGACA TAATCATAGG TCCTCTGACA CATAATTGCG

3490 3500 3510 3520 3530 3540
CTCTCTGATT AACGCCAGC GTTTCCGG TATCCAGATC CACAACCTTC GCTTCAAAAA

3550 3560 3570 3580 3590 3600
ATGGAACAAAC TTTACCGACC GCGCCCCGTT TATCATCCCC CTGGGGTGTATCAGAATAG

3610 3620 3630 3640 3650 3660
CTGATGTAGT CTCAGTGAGC CCATATCCTT GTCGTATCCC TGGAAGATGG AAGCGTTTGG

3670 3680 3690 3700 3710 3720
CAACCGCTTC CCCGACTTCTT TCGAAAGAG GTGCGCCCCC AGAAGCAATT TCGTGTAAAT

3730 3740 3750 3760 3770 3780
TAGATAAACATC GTATTTGTCA ATCAGAGTGC TTTGGCGAA GAATGAAAT AGGGTTGGTA

3790 3800 3810 3820 3830 3840
CTAGCAACGC ACTTTGAATT TTGTAATCCT GAAGGGATCG TAAAAACAGC TCTTCTCAA

3850 3860 3870 3880 3890 3900
ATCTATACAT TAAGACGACT CGAAATCCAC ATATCAAATA TCCGAGTGTATC GTAAACATTC

DNASIS

Molly

3910 3920 3930 3940 3950 3960
 CAAAACGTG ATGGAATGGA ACAACACTTAA AATCGCA GT ATCCGGAATG ATTGATTGC
 3970 3980 3990 4000 4010 4020
 CAAAAATAGG ATCTCTGGCA TGCGAGAAC TGACGCAGGC AGTTCTATGC GGAAGGGCCA
 4030 4040 4050 4060 4070 4080
 CACCCCTAGG TAACCCAGTA GATCCAGAGG AATTGTTTG TCACGATCAA AGGACTCTGG
 4090 4100 4110 4120 4130 4140
 TACAAAATCG TATTCACTAA AACCGGGAGG TAGATGAGAT GTGACGAACG TGTACATCGA
 4150 4160 4170 4180 4190 4200
 CTGAAATCCC TGGTAATCCG TTTTACAATC CATGATAATA ATTTCTGGA TTATTGGTAA
 4210 4220 4230 4240 4250 4260
 TTTTTTTGCA ACGTTCAAA TTTTTGCAA CCCCTTTTG GAAACAAACA CTACGGTAGG
 4270 4280 4290 4300 4310 4320
 CGAAATG TTCATACTGT TGAGCAATTAC ACGTCATTA TAAATGTCGT TCGCGGGCGC
 4330 4340 4350 4360 4370 4380
 AACTGCAACT CCGATAAATA ACGCGCCCAA CACCGGCATA AAGAATTGAA GAGAGTTTC
 4390 4400 4410 4420 4430 4440
 ACTGCATAAG ACGATTCTGT GATTTGTATT CAGCCCATAT CGTTTCATAG CTTCTGCCAA
 4450 4460 4470 4480 4490 4500
 CCGAACGGAC ATTCGAAGT ATTCCCGTGA CGTGATGTT ACCTCGATAT GTGCATCTGT
 4510 4520 4530 4540 4550 4560
 AAAAGGAATT GTTCCAGGAA CCAGGGCGTA TCTCTTCATA GCCTTATGCA GTTGCTCTCC
 4570 4580 4590 4600 4610 4620
 AGCGGTTCCA TCCTCTAGCT TTGCTTCTCA ATTTCTTATT TGATAATGA GAAAAAAAGG
 4630 4640 4650 4660 4670 4680
 TTTAAATT TTAACACCAA TTCAGTAGTT GATTGAGCAA ATGCCTTGCC AAAAAGGATG
 4690 4700 4710 4720 4730 4740
 CTTTAGAGAC AGTGTCTCT GCACAGATAA GGACAAACAT TATTGAGAGG GAGTACCCAG
 4750 4760 4770 4780 4790 4800
 AGCTGAGACT CCTAAGCCAG TGAGTGGCAC AGCATCCAGG GAGAAATATG CTTGTCATCA
 4810 4820 4830 4840 4850 4860
 CCGAACGCCTG ATTCCGTAGA GCCACACCCCT GGTAAGGGCC AATCTGCTCA CACAGGATAG
 4870 4880 4890 4900 4910 4920
 AGAGGGCAGG AGCCAGGGCA GAGCATATAA GGTGAGGTAG GATCAGTTGC TCCTCACATT
 4930 4940 4950 4960 4970 4980
 TGCTTCTGAC ATAGTTGTGT TGGGAGCTTG GATCGATCCA CCATGGGCTT CAATACCCCTG
 4990 5000 5010 5020 5030 5040
 ATTGACTGGA ACAGCTGTAG CCCTGAACAG CAGCGTGCAG TGCTGACGCG TCCGGCGATT
 5050 5060 5070 5080 5090 5100
 TCCGCCTCTG ACAGTATTAC CCGGACGGTC AGCGATATTG TGGATAATGT AAAAACGCGC
 5110 5120 5130 5140 5150 5160
 GGTGACGATG CCCTGCGTGA ATACAGCGCT AAATTTGATA AACAGAAAGT GACAGCGCTA
 5170 5180 5190 5200 5210 5220
 CGCGTCACCC CTGAAGAGAT CGCCGCCGCC GGGCGCGTC TGAGCGACGA ATTAAAACAG

DNASIS
Molly

5230 5240 5250 5260 5270 5280
 GCGATGACCG CTGCCGTCAA AAATATTGAA ACGTTCCATT CCGCGCAGAC GCTACCGCCT
 5290 5300 5310 5320 5330 5340
 GTAGATGTGG AAACCCAGCC AGGCCTGCCT TGCCAGCAGG TTACCGTCTC CGTCTCGTCT
 5350 5360 5370 5380 5390 5400
 GTCGGTCTGT ATATCCCCGG CGGCTCGGCT CCGCTCTTCT CAACGGTGCT GATGCTGGCG
 5410 5420 5430 5440 5450 5460
 ACGCCGGCGC GCATTGCGGG ATGCCAGAAG GTGGTTCTGT GCTCGCCGCC GCCCATCGCT
 5470 5480 5490 5500 5510 5520
 GATGAAATCC TCTATGCGGC GCAACTGTGT GGCGTGCAGG AAATCTTAA CGTCGGCGC
 5530 5540 5550 5560 5570 5580
 CGCGAGGCCA TTGCCGCTCT GGCCCTCGGC AGCGAGTCCG TACCGAAAGT GGATAAAATT
 5590 5600 5610 5620 5630 5640
 .TGGCCCCG GCAACGCCCT TGTAACCGAA GCCAACGTC AGGTAGGCCA GCGTCTCGAC
 5650 5660 5670 5680 5690 5700
 GGCGCGGCTA TCGATATGCC AGCCGGGGCG TCTGAAGTAC TGGTGATCGC AGACAGCGC
 5710 5720 5730 5740 5750 5760
 GCAACACCGG ATTCGTCGC TTCTGACCTG CTCTCCCAGG CTGAGCACGG CCCGGATTCC
 5770 5780 5790 5800 5810 5820
 CAGGTGATCC TGCTGACGCC TGATGCTGAC ATTGCCCGA AGGTGGCGGA GGCGGTAGAA
 5830 5840 5850 5860 5870 5880
 CGTCAACTGG CGGAACTGCC GCGCGCGGAC ACCGCCCGGC AGGCCCTGAG CGCCAGTCGT
 5890 5900 5910 5920 5930 5940
 CTGATTGTGA CCAAAGATTT AGCGCAGTGC GTGCCATCT CTAATCAGTA TGGGCCGGAA
 5950 5960 5970 5980 5990 6000
 ACTTAATCA TCCAGACCGC CAATGCGCGC GATTTGGTGG ATGCGATTAC CAGCGCAGGC
 6010 6020 6030 6040 6050 6060
 TCGGTATTTT TC GGCGACTG GTGCCGGAA TCCGCCGGTG ATTACGCTTC CGGAACCAAC
 6070 6080 6090 6100 6110 6120
 CATGTTTAC CGACCTATGG CTATACTGCT ACCTGTTCCA GCCTTGGGTT AGCGGATTTC
 6130 6140 6150 6160 6170 6180
 CAGAAACGGA TGACCGTTCA GGAACTGTCG AAAGCGGGCT TTTCCGCTCT GGCACTAACCC
 6190 6200 6210 6220 6230 6240
 ATTGAAACAT TGGCGCGGC AGAACGTCTG ACCGCCATA AAAATGCCGT GACCCCTGCGC
 6250 6260 6270 6280 6290 6300
 GTAAACGCC C TCAAGGAGCA AGCATGAGGC ACTGAAAACA CTCTCAGCGT CGCTGACTTA
 6310 6320 6330 6340 6350 6360
 GCCCGTGGAA ATGTCCGCAA CCTGGAGATC CAGACATGAT AAGATACATT GATGAGTTG
 6370 6380 6390 6400 6410 6420
 GACAAACAC AACTAGAATG CAGTGAAAAA AATGCTTTAT TTGTGAAATT TGTGATGCTA
 6430 6440 6450 6460 6470 6480
 TTGCTTTATT TGTAACCATT ATAAGCTGCA ATAAACAAAGT TAACAACAAAC AATTGCAATT
 6490 6500 6510 6520 6530 6540

DNASIS
Molly L.

ATTTTATGTT TCAGGTTCA GGGGAGGTGT GGGAGGTTT TAAAGCAAG TAAAACCTCT
 6550 6560 6570 6580 6590 -6600
 ACAAAATGTGG TATGGCTGAT TATGATCTCT AGGGCCGGCC CTCGACGGCG CGCCTCTAGA
 6610 6620 6630 6640 6650 6660
 GCAGTGTGGT TTTGCAAGAG GAAGCAAAAA GCCTCTCCAC CCAGGCCTGG AATGTTCCA
 6670 6680 6690 6700 6710 6720
 CCCAATGTGAG AGCAGTGTGG TTTTGCAGA GGAAGCAAAA AGCCTCTCCA CCCAGGCCTG
 6730 6740 6750 6760 6770 6780
 GAATGTTTCC ACCCAATGTC GAGCAAACCC CGCCAGCGT CTTGTCAATT GCGAATTGCA
 6790 6800 6810 6820 6830 6840
 ACACGCAGAT GCAGTCGGGG CGGCGCGGTC CCAGGTCCAC TTGCAATT AAGGTGACGC
 6850 6860 6870 6880 6890 6900
 CTGTGGCCCTC GAACACCGAG CGACCTGCA GCCAATATGG GATCGGCCAT TGAACAAGAT
 6910 6920 6930 6940 6950 6960
 GGATTGCACG CAGGTTCTCC GGCGCTTGG GTGGAGAGGC TATTGGCTA TGACTGGGCA
 6970 6980 6990 7000 7010 7020
 CAACAGACAA TCGGCTGCTC TGATGCCGCC GTGTTCCGGC TGTCAGCGA GGGCGCCCG
 7030 7040 7050 7060 7070 7080
 GTTCTTTTG TCAAGACCGA CCTGTCCGGT GCCCTGAATG AACTGCAGGT AAGTGCAGGC
 7090 7100 7110 7120 7130 7140
 GTCGATGGCC GAGGCGGCCT CGGCCTCTGC ATAAATAAAA AAAATTAGTC AGCCATGCT
 7150 7160 7170 7180 7190 7200
 GGGGCGGAGA ATGGGCGGAA CTGGGCGGAG TTAGGGCGG GATGGGCGGA GTTAGGGCG
 7210 7220 7230 7240 7250 7260
 ~GACTATGGT TGCTGACTAA TTGAGATGCA TGCTTGCACT ACTTCTGCCT GCTGGGAGC
 7270 7280 7290 7300 7310 7320
 CTGGGACTT TCCACACCTG GTTGCTGACT AATTGAGATG CATGTTTGC ATACTTCTGC
 7330 7340 7350 7360 7370 7380
 CTGCTGGGGA GCCTGGGAC TTTCCACACC CTAACGTACA CACATTCCAC AGAATTAATT
 7390 7400 7410 7420 7430 7440
 CCCCTAGTTA TTAATAGTAA TCAATTACGG GGTCAATTAGT TCATAGCCCA TATATGGAGT
 7450 7460 7470 7480 7490 7500
 TCCCGGTTAC ATAACCTACG GTAAATGGCC CGCCTGGCTG ACCGCCAAC GACCCCCGCC
 7510 7520 7530 7540 7550 7560
 CATTGACGTC AATAATGACG TATGTTCCA TAGTAACGCC AATAGGGACT TTCCATTGAC
 7570 7580 7590 7600 7610 7620
 GTCAATGGGT GGACTATTAA CGGTAAACTG CCCACTTGGC AGTACATCAA GTGTATCATA
 7630 7640 7650 7660 7670 7680
 TGCCAAGTAC GCCCCCTATT GACGTCAATG ACGGTAAATG GCCCGCCTGG CATTATGCC
 7690 7700 7710 7720 7730 7740
 AGTACATGAC CTTATGGAC TTTCTACTT GGCAGTACAT CTACGTATTA GTCATCGCTA
 7750 7760 7770 7780 7790 7800
 TTACCATGGT GATGCGGTTT TGGCAGTACA TCAATGGCG TGGATAGCGG TTTGACTCAC

DNASIS
Molly L.

7810 7820 7830 7840 7850 7860
 GGGGATTTC C AAGTCTCCAC CCCATTGACG TCAATGGGAG TTTGTTTGG CACCAAATC
 7870 7880 7890 7900 7910 7920
 AACGGGACTT TCCAAAATGT CGTAACAAT CCGCCCCATT GACGCAAATG GGCAGTAGGC
 7930 7940 7950 7960 7970 7980
 GTGTACGGTG GGAGGTCTAT ATAAGCAGAG CTGGGTACGT GAACCGTCAG ATGCCCTGGA
 7990 8000 8010 8020 8030 8040
 GACGCCATCA CAGATCTCTC ACTATGGATT TTCAGGTGCA GATTATCAGC TTCTGCTAA
 8050 8060 8070 8080 8090 8100
 TCAGTGCTTC AGTCATAATG TCCAGAGGAC AAATTGTTCT CTCCCAGTCT CCAGCAATCC
 8110 8120 8130 8140 8150 8160
 TGTCTGCATC TCCAGGGGAG AAGGTACCAA TGACTTGCAG GGCCAGCTCA AGTGTAAAGTT
 8170 8180 8190 8200 8210 8220
 ATCCACTG GTTCCAGCAG AAGCCAGGAT CCTCCCCAA ACCCTGGATT TATGCCACAT
 8230 8240 8250 8260 8270 8280
 CCAACCTGGC TTCTGGAGTC CCTGTTCGCT TCAGTGGCAG TGGGTCTGGG ACTTCTTACT
 8290 8300 8310 8320 8330 8340
 CTCTCACAAAT CAGCAGAGTG GAGGCTGAAG ATGCTGCCAC TTATTACTGC CAGCAGTGG
 8350 8360 8370 8380 8390 8400
 CTAGTAACCC ACCCACGTTG GGAGGGGGGA CCAAGCTGGA AATCAAACGT ACGGTGGCTG
 8410 8420 8430 8440 8450 8460
 CACCATCTGT CTTCATCTTC CCGCCATCTG ATGAGCAGTT GAAATCTGGA ACTGCCCTTG
 8470 8480 8490 8500 8510 8520
 TTGTGTGCCT GCTGAATAAC TTCTATCCCA GAGAGGCCAA AGTACAGTGG AAGGTGGATA
 8530 8540 8550 8560 8570 8580
 GCCCTCCA ATCGGGTAAC TCCCAGGAGA GTGTACAGA GCAGGACAGC AAGGACAGCA
 8590 8600 8610 8620 8630 8640
 CCTACAGCCT CAGCAGCACC CTGACGCTGA GCAAAGCAGA CTACGAGAAA CACAAAGTCT
 8650 8660 8670 8680 8690 8700
 ACGCCTGCAG AGTCACCCAT CAGGGCCTGA GCTCGCCGT CACAAAGAGC TTCAACAGGG
 8710 8720 8730 8740 8750 8760
 GAGAGTGTG AATTCAAGATC CGTTAACGGT TACCAACTAC CTAGACTGGA TTCTGTACAA
 8770 8780 8790 8800 8810 8820
 CATGCGGCCG TGATATCTAC GTATGATCAG CCTCGACTGT GCCTTCTAGT TGCCAGCCAT
 8830 8840 8850 8860 8870 8880
 CTGTTGTTG CCCCTCCCCC GTGCCTTCCCT TGACCCCTGGA AGGTGCCACT CCCACTGTCC
 8890 8900 8910 8920 8930 8940
 TTTCTAATA AAATGAGGAA ATTGCATCGC ATTGTCTGAG TAGGTGTCA TCTATTCTGG
 8950 8960 8970 8980 8990 9000
 GGGGTGGGGT GGGGCAGGAC AGCAAGGGGG AGGATTGGGA AGACAATAGC AGGCATGCTG
 9010 9020 9030 9040 9050 9060
 GGGATGCGGT GGGCTCTATG GAACCAGCTG GGGCTCGACA GCTATGCCAA GTACGCCCT
 9070 9080 9090 9100 9110 9120
 TATTGACGTC AATGACGGTA AATGGCCCGC CTGGCATTAT GCCCAGTACA TGACCTTATG

DNASIS
Molly

9130	9140	9150	9160	9170	9180
GGACTTTCT	ACTTGGCAGT	ACATCTACGT	ATTAGTCATC	GCTATTACCA	TGGTGATGCG
9190	9200	9210	9220	9230	9240
GTTTTGGCAG	TACATCAATG	GGCGTGGATA	GCGGTTTGAC	TCACGGGAT	TTCCAAGTCT
9250	9260	9270	9280	9290	9300
CCACCCCCATT	GACGTCAATG	GGAGTTTGT	TTGGCACCAA	AATCAACGGG	ACTTTCCAAA
9310	9320	9330	9340	9350	9360
ATGTCGTAAC	AACTCCGCC	CATTGACGCA	AATGGGCGGT	AGGCCTGTAC	GGTGGGAGGT
9370	9380	9390	9400	9410	9420
CTATATAAGC	AGAGCTGGGT	ACGTCTCAC	ATTCAGTGAT	CAGCACTGAA	CACAGACCCG
9430	9440	9450	9460	9470	9480
TCGACATGGG	TTGGAGCCTC	ATCTTGCTCT	TCCTTGTGCG	TGTTGCTACG	CGTGTCTGT
9490	9500	9510	9520	9530	9540
CCCAGGTACA	ACTGCAGCAG	CCTGGGGCTG	AGCTGGTGA	GCCTGGGGCC	TCAGTGAAGA
9550	9560	9570	9580	9590	9600
TGTCTGCAA	GGCTTCTGGC	TACACATTAA	CCAGTTACAA	TATGCACTGG	GTAAAACAGA
9610	9620	9630	9640	9650	9660
CACCTGGTCG	GGGCCTGGAA	TGGATTGGAG	CTATTTATCC	CGGAAATGGT	GATACTTCCT
9670	9680	9690	9700	9710	9720
ACAATCAGAA	GTTCAAAGGC	AAGGCCACAT	TGACTGCAGA	CAAATCCTCC	AGCACAGCCT
9730	9740	9750	9760	9770	9780
ACATGCAGCT	CAGCAGCCTG	ACATCTGAGG	ACTCTGCGGT	CTATTACTGT	GCAAGATCGA
9790	9800	9810	9820	9830	9840
CTTACTACGG	CGGTGACTGG	TACTTCAATG	TCTGGGGCGC	AGGGACCACG	GTCACCGTCT
9850	9860	9870	9880	9890	9900
CTGCAGCTAG	CACCAAGGGC	CCATCGGTCT	TCCCCCTGGC	ACCCCTCCTCC	AAGAGCACCT
9910	9920	9930	9940	9950	9960
CTGGGGGCAC	AGCGGCCCTG	GGCTGCCTGG	TCAAGGACTA	CTTCCCCGAA	CCGGTGACGG
9970	9980	9990	10000	10010	10020
TGTCTGGAA	CTCAGGCGCC	CTGACCAGCG	GCGTGCACAC	CTTCCCGGT	GTCTTACAGT
10030	10040	10050	10060	10070	10080
CCTCAGGACT	CTACTCCCTC	AGCAGCGTGG	TGACCGTGCC	CTCCAGCAGC	TTGGGCACCC
10090	10100	10110	10120	10130	10140
AGACCTACAT	CTGCAACGTG	AATCACAAGC	CCAGCAACAC	CAAGGTGGAC	AAGAAAGCAG
10150	10160	10170	10180	10190	10200
AGCCCAAATC	TTGTGACAAA	ACTCACACAT	GCCCACCGTG	CCCAGCACCT	GAACTCCTGG
10210	10220	10230	10240	10250	10260
GGGGACCGTC	AGTCTCCTC	TTCCCCCAA	AACCCAAGGA	CACCCCTCATG	ATCTCCCGGA
10270	10280	10290	10300	10310	10320
CCCCTGAGGT	CACATGCCTG	GTGGTGGACG	TGAGGCCACGA	AGACCCCTGAG	GTCAAGTTCA
10330	10340	10350	10360	10370	10380
ACTGGTACGT	GGACGGCGTG	GAGGTGCATA	ATGCCAAGAC	AAAGCCGGG	GAGGAGCAGT
10390	10400	10410	10420	10430	10440

DNASIS
Molly

ACAAACAGCAC GTACCGTGTG GTCAGCGTCC TCACCGTCT GCACCAGGAC TGGCTGAATG

10450 10460 10470 10480 10490 10500
GCAAGGAGTA CAAGTGCAAG GTCTCCAACA AAGCCCTCCC AGCCCCCATC GAGAAAACCA

10510 10520 10530 10540 10550 10560
TCTCCAAAGC CAAAGGGCAG CCCCAGAAC CACAGGTGTA CACCCCTGCC CCATCCCAGGG

10570 10580 10590 10600 10610 10620
ATGAGCTGAC CAAGAACCAAG GTCAGCCTGA CCTGCCTGGT CAAAGGCTTC TATCCCAGCG

10630 10640 10650 10660 10670 10680
ACATCGCCGT GGAGTGGGAG AGCAATGGGC AGCCGGAGAA CAACTACAAG ACCACGCCTC

10690 10700 10710 10720 10730 10740
CCGTGCTGGA CTCCGACGGC TCCTTCTTCC TCTACAGCAA GCTCACCGTG GACAAGAGCA

10750 10760 10770 10780 10790 10800
GTGGCAGCA GGGGAACGTC TTCTCATGCT CCGTGATGCA TGAGGCTCTG CACAACACT

10810 10820 10830 10840 10850 10860
ACACGCAGAA GAGCCTCTCC CTGTCTCCGG GTAAATGAGG ATCCGTTAAC GGTTACCAAC

10870 10880 10890 10900 10910 10920
TACCTAGACT GGATTGCGTGA CAACATGCGG CCGTGATATC TACGTATGAT CAGCCTCGAC

10930 10940 10950 10960 10970 10980
TGTGCCTTCT AGTTGCCAGC CATCTGTTGT TTGCCCTC CCCGTGCCTT CCTTGACCCCT

10990 11000 11010 11020 11030 11040
GGAAGGTGCC ACTCCCAGT TCCTTCTCTA ATAAAATGAG GAAATTGCAT CGCATTGTCT

11050 11060 11070 11080 11090 11100
GAGTAGGTGT CATTCTATTC TGGGGGGTGG GGTGGGGCAG GACAGCAAGG GGGAGGGATTG

11110 11120 11130 11140 11150 11160
GAAGACAAT AGCAGGCATG CTGGGGATGC GGTGGGCTCT ATGGAACCAAG CTGGGGCTCG

11170 11180 11190 11200 11210 11220
ACAGCAACGC TAGGTCGAGG CCGCTACTAA CTCTCTCTC CCTCCTTTT CCTGCAGGAC

11230 11240 11250 11260 11270 11280
GAGGCAGCGC GGCTATCGTG GCTGGCCACG ACGGGGCGTTC CTTGCGCAGC TGTGCTCGAC

11290 11300 11310 11320 11330 11340
GTTGTCAGT AAGCGGGAAAG GGACTGGCTG CTATTGGCG AAGTGCCGGG GCAGGATCTC

11350 11360 11370 11380 11390 11400
CTGTCATCTC ACCTTGCTCC TGCCGAGAAA GTATCCATCA TGGCTGATGC AATGCGCGG

11410 11420 11430 11440 11450 11460
CTGCATAACGC TTGATCCGGC TACCTGCCA TTGCGACCAAC AAGCGAAACA TCGCATCGAG

11470 11480 11490 11500 11510 11520
CGAGCACGTA CTCGGATGGA AGCCGGTCTT GTCGATCAGG ATGATCTGGA CGAAGAGCAT

11530 11540 11550 11560 11570 11580
CAGGGGGCTCG CGCCAGCCGA ACTGTTGCC AGGTAAGTGA GCTCCAATTC AAGCTTCCCTA

11590 11600 11610 11620 11630 11640
GGCGGGCCAG CTAGTAGCTT TGCTTCTCAA TTTCTTATTT GCATAATGAG AAAAAAAGGA

11650 11660 11670 11680 11690 11700
AAATTAATTT TAACACCAAT TCAGTAGTTG ATTGAGCAAA TGCCTTGCCA AAAAGGATGC

DNASIS
Molly

11710	11720	11730	11740	11750	11760
TTTAGAGACA GTGTTCTCTG CACAGATAAG GACAAACATT ATTCAAGAGGG AGTACCCAGA					
11770	11780	11790	11800	11810	11820
GCTGAGACTC CTAAGCCAGT GAGTGGACA GCATCCAGGG AGAAATATGC TTGTCATCAC					
11830	11840	11850	11860	11870	11880
CGAACGCTGA TTCCGTAGAG CCACACCCCTG GTAAGGGCCA ATCTGCTCAC ACAGGATAGA					
11890	11900	11910	11920	11930	11940
GAGGGCAGGA GCCAGGGCAG AGCATATAAG GTGAGGTTAGG ATCAGTTGCT CCTCACATT					
11950	11960	11970	11980	11990	12000
GCTTCTGACA TAGTTGTGTT GGGAGCTTGG ATAGCTTGGG GGGGGGACAG CTCAGGGCTG					
12010	12020	12030	12040	12050	12060
CGATTTCGCG CCAAACTTGA CGGCAATCCT AGCGTGAAGG CTGGTAGGAT TTTATCCCCG					
12070	12080	12090	12100	12110	12120
GCCATCAT GGTTGACCA TTGAAC TGCA TCGTCGCCGT GTCCAAAAT ATGGGGATTG					
12130	12140	12150	12160	12170	12180
GCAAGAACGG AGACCTACCC TGGCCTCCGC TCAGGAACGA GTTCAAGTAC TTCAAAGAA					
12190	12200	12210	12220	12230	12240
TGACCACAAAC CTCTTCAGTG GAAGGTAAAC AGAATCTGGT GATTATGGGT AGGAAAACCT					
12250	12260	12270	12280	12290	12300
GGTTCTCCAT TCCTGAGAAG AATCGACCTT TAAAGGACAG AATTAATATA GTTCTCAGTA					
12310	12320	12330	12340	12350	12360
GAGAACTCAA AGAACCAAGA CGAGGAGCTC ATTTTCTTGC CAAAAGTTG GATGATGCCT					
12370	12380	12390	12400	12410	12420
TAAGACTTAT TGAACAAACCG GAATTGGCAA GTAAAGTAGA CATGGTTGG ATAGTCGGAG					
12430	12440	12450	12460	12470	12480
AGTTCTGT TTACCAAGGAA GCCATGAATC AACCAGGCCA CCTCAGACTC TTTGTGACAA					
12490	12500	12510	12520	12530	12540
GGATCATGCA GGAATTGAA AGTGACACGT TTTTCCCAGA AATTGATTG GGGAAATATA					
12550	12560	12570	12580	12590	12600
AACTTCTCCC AGAATACCCA GGCGTCTCT CTGAGGTCCA GGAGGAAAAA GGCATCAAGT					
12610	12620	12630	12640	12650	12660
ATAAGTTGA AGTCTACGAG AAGAAAGACT AACAGGAAGA TGCTTCAAG TTCTCTGCTC					
12670	12680	12690	12700	12710	12720
CCCTCCTAAA GCTATGCATT TTTATAAGAC CATGGGACTT TTGCTGGCTT TAGATCAGCC					
12730	12740	12750	12760	12770	12780
TCGACTGTGC CTTCTAGTTG CCAGCCATCT GTTGTGGGCC CCTCCCCGT GCCTTCTTG					
12790	12800	12810	12820	12830	12840
ACCCCTGGAAG GTGCCACTCC CACTGTCCTT TCCTAATAAA ATGAGGAAT TGCACTGCAT					
12850	12860	12870	12880	12890	12900
TGTCTGAGTA GGTGTCATTTC TATTCTGGGG GGTGGGGTGG GGCAGGACAG CAAGGGGGAG					
12910	12920	12930	12940	12950	12960
GATTGGGAAG ACAATAGCAG GCATGCTGGG GATGCGGTGG GCTCTATGGC TTCTGAGGCG					
12970	12980	12990	13000	13010	13020
GAAAGAACCA GCTGGGGCTC GAAGCGGCCG CCCATTTGCG TGGTGGTCAG ATGCGGGATG					

DNASIS
Molly L.

13030 13040 13050 13060 13070 13080
 CGGTGGGACG CGGCAGGGAG CGTCACACTG AGGTTTCCG CCAGACGCCA CTGCTGCCAG

 13090 13100 13110 13120 13130 13140
 GCGCTGATGT GCCCGGCTTC TGACCATGCCG GTCCGGTTCG GTTGCAC TAC GCGTACTGTG

 13150 13160 13170 13180 13190 13200
 AGCCAGAGTT GCCCGGCGCT CTCCGGCTGC GGTAGTTCA GAGTTCAAT CAACTGTTA

 13210 13220 13230 13240 13250 13260
 CCTTGTTGGAG CGACATCCAG AGGCAC TTCA CCGCTTGCCA GCGGCTTACCC ATCCAGCGCC

 13270 13280 13290 13300 13310 13320
 ACCATCCAGT GCAGGAGCTC GTTATCGCTA TGACGGAACA GGTATTGCT GGTCACTTCG

 13330 13340 13350 13360 13370 13380
 ATGGTTTGC CCGATAAACG GAACTGGAAA AACTGCTGCT GGTGTTTGC TTCCGTCAAGC

 13390 13400 13410 13420 13430 13440
 GGATGCG GCGTGGGTC GGCAAAGACC AGACCGTTCA TACAGAACTG GCGATCGTTC

 13450 13460 13470 13480 13490 13500
 GGC GTATCGC CAAATCACCC GCGTAAGCC GACCACGGGT TGCCGTTTTC ATCATATTTA

 13510 13520 13530 13540 13550 13560
 ATCAGCGACT GATCCACCCCA GTCCCAGACG AAGCCGCCCT GTAAACGGGG ATACTGACGA

 13570 13580 13590 13600 13610 13620
 AACGCCCTGCC AGTATTTAGC GAAACGCCA AGACTGTTAC CCATCGCGTG GCGTATTTCG

 13630 13640 13650 13660 13670 13680
 CAAAGGATCA GCGGGCGCGT CTCTCCAGGT AGCGAAAGCC ATTGTTTGAT GGACCATTTC

 13690 13700 13710 13720 13730 13740
 GGCACAGCCG GGAAGGGCTG GTCTTCATCC ACGCGCGCGT ACATCGGGCA AATAATATCG

 13750 13760 13770 13780 13790 13800
 GGC CGTGG TGTGGCTCC GCGCCCTCA TACTGCACCG GCGGGGAAGG ATCGACAGAT

 13810 13820 13830 13840 13850 13860
 TTGATCCAGC GATACAGCGC GTCGTGATTA GCGCCGTGGC CTGATTCACTT CCCCAGCGAC

 13870 13880 13890 13900 13910 13920
 CAGATGATCA CACTCGGGTG ATTACGATCG CGCTGCACCA TTCGCGTTAC GCGTTCGCTC

 13930 13940 13950 13960 13970 13980
 ATCGCCGGTA GCCAGCGCGG ATCATCGGTC AGACGATTCA TTGGCACCAT GCGTGGGTT

 13990 14000 14010 14020 14030 14040
 TCAATATTGG CTTCATCCAC CACATACAGG CCGTAGCGGT CGCACAGCGT GTACCACAGC

 14050 14060 14070 14080 14090 14100
 GGATGGTTCG GATAATGCGA ACAGCGCACG GCGTTAAAGT TGTCTGCTT CATCAGCGAGG

 14110 14120 14130 14140 14150 14160
 ATATCCTGCA CCATCGTCTG CTCATCCATG ACCTGACCAT GCAGAGGATG ATGCTCGTGA

 14170 14180 14190 14200 14210 14220
 CGGTTAACGC CTCGAATCAG CAACGGCTTG CCGTTCA GCGAGCAGACC ATTTCAATC

 14230 14240 14250 14260 14270 14280
 CGCACCTCGC GGAAACCGAC ATCGCAGGCT TCTGCTTCAA TCAGCGTGCC GTCGGCGGTG

 14290 14300 14310 14320 14330 14340

DNASIS
Molly L.

TGCAGTTCAA CCACCGCACG ATAGAGATTG GGGATTCGG CGCTCCACAG TTTCGGTTT
 14350 14360 14370 14380 14390 14400
 TCGACGTTCA GACGTAGTGT GACCGCATCG GCATAACCAC CACGCTCATC GATAATTCA
 14410 14420 14430 14440 14450 14460
 CCGCCGAAAG GCGCGGTGCC GCTGGCGACC TGCGTTTCAC CCTGCCATAA AGAAACTGTT
 14470 14480 14490 14500 14510 14520
 ACCCGTAGGT AGTCACGCAA CTCGCCGAC ATCTGAACCTT CAGCCTCCAG TACAGCCGG
 14530 14540 14550 14560 14570 14580
 CTGAAATCAT CATTAAAGCG AGTGGCAACA TGGAATCGC TGATTTGTGT AGTCGGTTA
 14590 14600 14610 14620 14630 14640
 TGCAGCAACG AGACGTACG GAAAATGCCG CTCATCCGCC ACATATCCCTG ATCTTCCAGA
 14650 14660 14670 14680 14690 14700
 TAACTGCCGT CACTCCAGCG CAGCACCATC ACCGCGAGGC GGTTTTCTCC GGCCTCGTAAA
 14710 14720 14730 14740 14750 14760
 AATGCGCTCA GGTCAAATTG AGACGGCAAA CGACTGTCTT GGCCGTAACC GACCCAGCGC
 14770 14780 14790 14800 14810 14820
 CCGTTGCACC ACAGATGAAA CGCCGAGTTA ACGCCATCAA AAATAATTG CGTCTGGCCT
 14830 14840 14850 14860 14870 14880
 TCCTGTAGCC AGCTTTCATC AACATTAAT GTGAGCGAGT AACAAACCGT CGGATTCTCC
 14890 14900 14910 14920 14930 14940
 GTGGGAACAA ACGGCGGATT GACCGTAATG GGATAGGTGA CGTTGGTGTA GATGGCGCA
 14950 14960 14970 14980 14990 15000
 TCGTAACCGT GCATCTGCCA GTTGAGGGG ACGACGACAG TATCGGCCTC AGGAAGATCG
 15010 15020 15030 15040 15050 15060
 CACTCCAGCC AGCTTCCGG CACCGCTTCT GGTGCCGGAA ACCAGGCAA GCGCCATTG
 15070 15080 15090 15100 15110 15120
 CCATTCAAGGC TGCGCAACTG TTGGGAAGGG CGATCGGTGC GGGCCTCTTC GCTATTACGC
 15130 15140 15150 15160 15170 15180
 CAGCTGGCGA AAGGGGGATG TGCTGCAAGG CGATTAAGTT GGGTAACGCC AGGGTTTTCC
 15190 15200 15210 15220 15230 15240
 CAGTCACGAC GTTGTAAAAC GACTTAATCC GTCGAGGGGC TGCCCTGAAG CAGACGACCT
 15250 15260 15270 15280 15290 15300
 TCCGTTGTGC AGCCAGCGGC GCCTGCGCCG GTGCCACAA TCGTGCACGA ACAAACTAAA
 15310 15320 15330 15340 15350 15360
 CCAGAACAAA TTATACCGGC GGCACCGCCG CCACCACTT CTCCCGTGCCT TAACATTCCA
 15370 15380 15390 15400 15410 15420
 GCGCCTCCAC CACCAACCAC ACCATCGATG TCTGAATTGC CGCCCGCTCC ACCAATGCCG
 15430 15440 15450 15460 15470 15480
 ACGGAACCTC AACCCGCTGC ACCTTTAGAC GACAGACAA AATTGTTGGA AGCTATTAGA
 15490 15500 15510 15520 15530 15540
 AACGAAAAAA ATCGCACTCG TCTCAGACCG GTCAAACAA AAACGGCGCC CGAAACCAAGT
 15550 15560 15570 15580 15590 15600
 ACAATAGTTG AGGTGCCGAC TGTGTTGCCT AAAGAGACAT TTGAGCCTAA ACCGCCGTCT

DNASIS
Molly L

15610 15620 15630 15640 15650 15660
 GCATCACCGC CACCACTCC GCCTCCGCT CCGCCGCCAG CCCCACCTGC GCCTCCACCG
 15670 15680 15690 15700 15710 15720
 ATGGTAGATT TATCATCAGC TCCACCACCG CCGCATTAG TAGATTTGCC GTCTGAAATG
 15730 15740 15750 15760 15770 15780
 TTACCAACCGC CTGCACCATC GCTTCTAAC GTGTTGTCTG AATTAAAATC GGGCACAGTT
 15790 15800 15810 15820 15830 15840
 AGATTGAAAC CCGCCCAAAA ACGCCCGCAA TCAGAAATAA TTCCAAAAAG CTCAACTACA
 15850 15860 15870 15880 15890 15900
 AATTTGATCG CGGACGTGTT AGCCGACACA ATTAATAGGC GTCTGTGGC TATGGCAAAA
 15910 15920 15930 15940 15950 15960
 TCGTCTTCGG AAGCAACTTC TAACGACGAG GGTTGGGACG ACGACGATAA TCGGCCTAAT
 15970 15980 15990 16000 16010 16020
 AGCTAACAC CGCCCGATGT TAAATATGTC CAAGCTACTA GTGGTACCGC TTGGCAGAAC
 16030 16040 16050 16060 16070 16080
 ATATCCATCG CGTCCGCCAT CTCCAGCAGC CGCACCGGGC GCATCTCGGG CAGCGTTGGG
 16090 16100 16110 16120 16130 16140
 TCCTGGCAC GGGTGCACAT GATCGTCTC CTGTGTTGA GGACCCGGCT AGGCTGGCGG
 16150 16160 16170 16180 16190 16200
 GGTTGCCTTA CTGGTTAGCA GAATGAATCA CCGATACGCG AGCGAACGTG AAGCGACTGC
 16210 16220 16230 16240 16250 16260
 TGCTGCAAAA CGTCTGCAC CTGAGCAACA ACATGAATGG TCTTCGGTTT CCTGTCTCG
 16270 16280 16290 16300 16310 16320
 TAAAGTCTGG AAACCGGAA GTCAGCGCC TGCAACATTA TGTTCCGGAT CTGCATCGCA
 16330 16340 16350 16360 16370 16380
 GATGCTGCT GGCTACCTTG TGGAACACCT ACATCTGTAT TAACGAAGCG CTGGCATTGA
 16390 16400 16410 16420 16430 16440
 CCCTGAGTGA TTTTCTCTG GTCCCGCCGC ATCCATACCG CCAGTTGTTT ACCCTCACAA
 16450 16460 16470 16480 16490 16500
 CGTTCCAGTA ACCGGGCATG TTCATCATCA GTAAACCGTA TCGTGAGCAT CCTCTCTCGT
 16510 16520 16530 16540 16550 16560
 TTCAATCGGTA TCATTACCCC CATGAACAGA AATCCCCCTT ACACGGAGGC ATCAAGTGA
 16570 16580 16590 16600 16610 16620
 AACAGGAAA AAACCGCCCT TAACATGGCC CGCTTATCA GAAGCCAGAC ATTAACGCTT
 16630 16640 16650 16660 16670 16680
 CTGGAGAAAAC TCAACGAGCT GGACGCGGAT GAACAGGCAG ACATCTGTGA ATCGCTTCAC
 16690 16700 16710 16720 16730 16740
 GACCACGCTG ATGAGCTTTA CCGCAGCTGC CTCGCGCGTT TCGGTGATGA CGGTGAAAAC
 16750 16760 16770 16780 16790 16800
 CTCTGACACA TGCAGCTCCC GGAGACGGTC ACAGCTTGTC TGTAAGCGGA TGCCGGGAGC
 16810 16820 16830 16840 16850 16860
 AGACAAGCCC GTCAGGGCGC GTCAGCGGGT GTTGGCGGGT GTCGGGGCGC AGCCATGACC
 16870 16880 16890 16900 16910 16920
 CAGTCACGTA GCGATAGCGG AGTGTATACT GGCTTAACCA TGCGGCATCA GAGCAGATTG

DNASIS
Molly Park

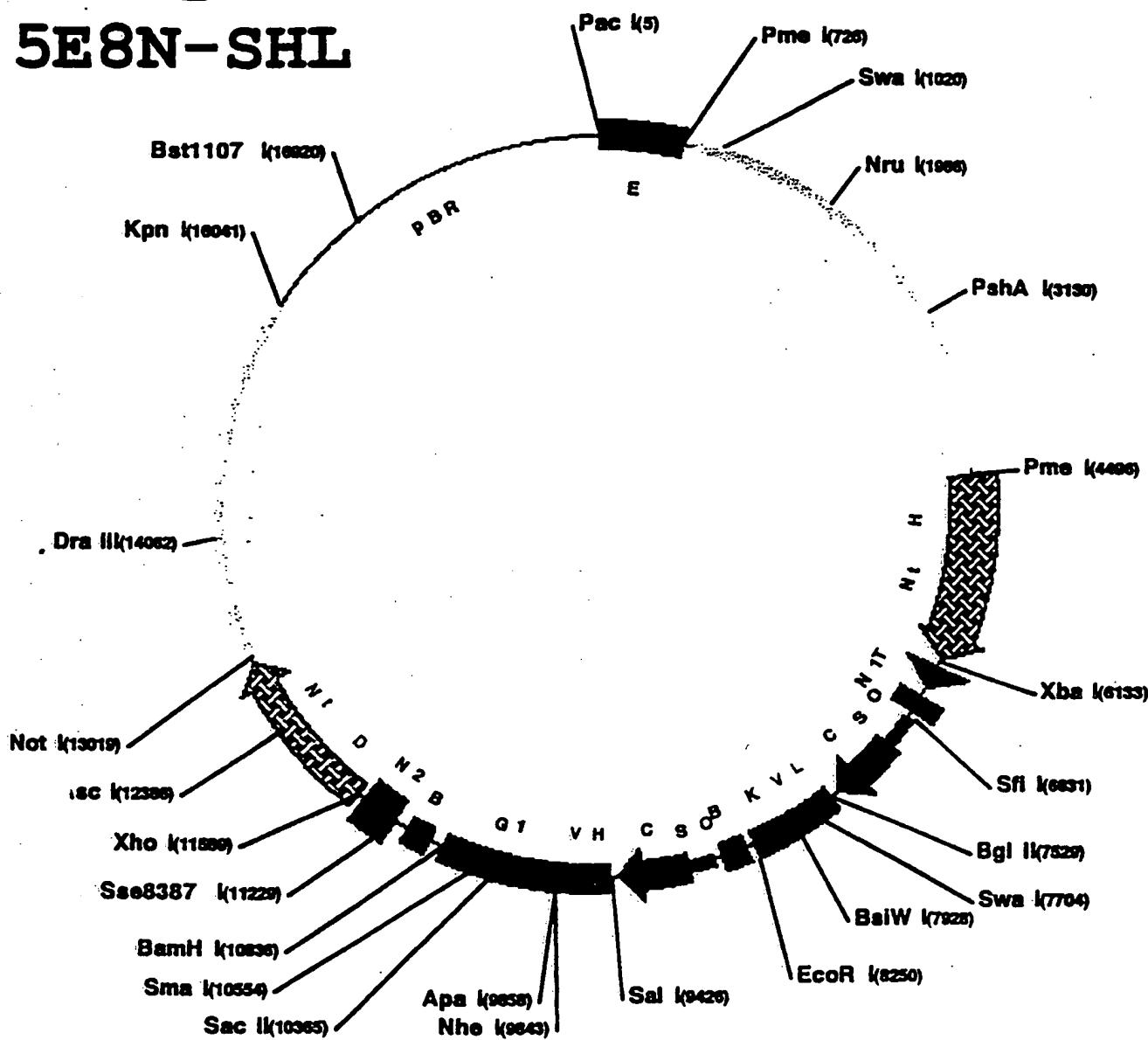
16930	16940	16950	16960	16970	16980
TACTGAGAGT GCACCATATG CGGTGTGAAA TACCGCACAG ATGCCTAAGG AGAAAATACC					
16990	17000	17010	17020	17030	17040
GCATCAGGCG CTCTCCGCT TCCTCGCTCA CTGACTCGCT GCGCTGGTC GTTCGGCTGC					
17050	17060	17070	17080	17090	17100
GGCGAGGGT ATCAGCTCAC TCAAAGGCGG TAATACGGTT ATCCACAGAA TCAGGGGATA					
17110	17120	17130	17140	17150	17160
ACGCAGGAAA GAACATGTGA GCAAAAGGCC AGCAAAAGGC CAGGAACCGT AAAAAGGCCG					
17170	17180	17190	17200	17210	17220
CGTTGCTGGC GTTTTCCAT AGGCTCCGCC CCCCTGACGA GCATCACAAA AATCGACGCT					
17230	17240	17250	17260	17270	17280
CAAGTCAGAG GTGGCGAAC CCGACAGGAC TATAAAGATA CCAGGCGTTT CCCCCCTGGAA					
17290	17300	17310	17320	17330	17340
GCTCCCTCGT GCGCTCTCCT GTTCCGACCC TGCCGCTTAC CGGATACTG TCCGCCCTTC					
17350	17360	17370	17380	17390	17400
TCCCTTCGGG AAGCGTGGCG CTTCTCATA GCTCACGCTG TAGGTATCTC AGTCGGTGT					
17410	17420	17430	17440	17450	17460
AGGTCGTTCG CTCCAAGCTG GGCTGTGTGC ACGAACCCCC CGTTCAGCCC GACCGCTGCG					
17470	17480	17490	17500	17510	17520
CCTTATCCGG TAACTATCGT CTTGAGTCCA ACCCGGTAAG ACACGACTTA TCGCCACTGG					
17530	17540	17550	17560	17570	17580
CAGCAGCCAC TGGTAACAGG ATTACGAGAG CGAGGTATGT AGGCGGTGCT ACAGAGTTCT					
17590	17600	17610	17620	17630	17640
TGAAGTGGTG GCCTAACTAC GGCTACACTA GAAGGACAGT ATTTGGTATC TGCCTCTGC					
17650	17660	17670	17680	17690	17700
TGAAGCCAGT TACCTTCGGA AAAAGAGTTG GTAGCTTTG ATCCGGCAA CAAACCACCG					
17710	17720	17730	17740	17750	17760
CTGGTAGCCGG TGGTTTTTTT GTTGCAAGC AGCAGATTAC GCGCAGAAAA AAAGGATCTC					
17770	17780	17790	17800	17810	17820
AAGAAGATCC TTGATCTTT TCTACGGGGT CTGACGCTCA GTGGAACGAA AACTCACGTT					
17830	17840	17850	17860	17870	17880
AAGGGATTTT GGTCACTGAGA TTATCAAAAA GGATCTTCAC CTAGATCCTT TAAATTAAA					
17890	17900	17910	17920	17930	17940
AATGAAGTTT TAAATCAATC TAAAGTATAT ATGAGTAAAC TTGGTCTGAC AGTTACCAAT					
17950	17960	17970	17980	17990	18000
GCTTAATCAG TGAGGCACCT ATCTCAGCGA TCTGTCTATT TCGTTCATCC ATAGTTGCCT					
18010	18020	18030	18040	18050	18060
GACTCCCCGT CGTGTAGATA ACTACGATAC GGGAGGGCTT ACCATCTGGC CCCAGTGCTG					
18070	18080	18090	18100	18110	18120
CAATGATACC GCGAGACCA CGCTCACCGG CTCCAGATTT ATCAGCAATA AACCAAGCCAG					
18130	18140	18150	18160	18170	18180
CCGGAAGGGC CGAGCGAGA AGTGGTCCTG CAACTTTATC CGCCTCCATC CAGTCTATTA					
18190	18200	18210	18220	18230	18240

DNASIS
Molly Lark

ATTGTTGCCG GGAAGCTAGA GTAAGTAGTT CGCCAGTTAA TAGTTTGCAC AACGTTGTTG
18250 18260 18270 18280 18290 18300
CCATTGCTGC AGGCATCGTG GTGTCACGCT CGTCGTTGG TATGGCTTCA TTCAGCTCCG
18310 18320 18330 18340 18350 18360
GTTCCCAACG ATCAAGGCAGA GTTACATGAT CCCCCATGTT GTGAAAAAA GCGGTTAGCT
18370 18380 18390 18400 18410 18420
CCTTCGGTCC TCCGATCGTT GTCAGAAGTA AGTTGGCCGC AGTGTATCA CTCATGGTTA
18430 18440 18450 18460 18470 18480
TGGCAGCACT GCATAATTCT CTTACTGTCA TGCCATCCGT AAGATGCTTT TCTGTGACTG
18490 18500 18510 18520 18530 18540
GTGAGTACTC AACCAAGTCA TTCTGAGAAT AGTGTATGCG GCGACCGAGT TGCTCTGCC
18550 18560 18570 18580 18590 18600
GGCGTCAAC ACGGGATAAT ACCGCGCAC ATAGCAGAAC TTTAAAAGTG CTCATCATTG
18610 18620 18630 18640 18650 18660
GAAAACGTTTC TTGGGGCGA AAACTCTCAA GGATCTTACC GCTGTTGAGA TCCAGTTCGA
18670 18680 18690 18700 18710 18720
TGTAACCCAC TCGTGCACCC AACTGATCTT CAGCATCTT TACTTTCACC AGCGTTCTG
18730 18740 18750 18760 18770 18780
GGTGAGCAAA AACAGGAAGG CAAATGCCG CAAAAAAGGG AATAAGGGCG ACACGGAAAT
18790 18800 18810 18820 18830 18840
GTTGAATACT CATACTCTTC CTTTTCAAT ATTATTGAAG CATTATCAG GTTTATTGTC
18850 18860 18870 18880 18890 18900
TCATGAGCGG ATACATATTT GAATGTATTT AGAAAAATAA ACAAAATAGGG GTTCCGCGCA
18910 18920 18930 18940 18950 18960
-TTTCCCCG AAAAGTCCA CCTGACGTCT AAGAAAACCAT TATTATCATG ACATTAACCT
18970 18980 18990 19000 19010 19020
ATAAAAATAG GCGTATCACG AGGCCCTTC GTCTTCAAGA A.....

Mandy + 5E8N-SHL

FIGURE 9



Nt D = Inactive Dihydrofolate reductase
 E = CMV and SV40 enhancers

SO = SV40 Origin of replication

Nt H = Inactive *Salmonella* Histidinol Dehydrogenase

T = Herpes Simplex thymidine kinase promoter and polyoma enhancer

C = Cytomegalovirus promoter/enhancer B = Bovine growth hormone polyadenylation

N1 = Neomycin phosphotransferase exon 1 M2 = Neomycin phosphotransferase exon 2

K = Human kappa constant G1 = Human Gamma 1 constant

VL = Variable light chain anti-CD23 primate 5E8 and leader

VH = Variable heavy chain anti-CD23 primate 5E8N- and leader

Mandy cut Xba I and ligated to Xba I Xba I fragment from XKG1+CD23 5E8N-SHL

FIGURE 10

DNASIS
Mandy E8N-SHL

10 20 30 40 50 60
 TTAATTAAGG GGC GGAGAAT GGG CGGA ACT GGG CGG AGTT AGGGG CGGG A TG GCG GAGT
 70 80 90 100 110 120
 TAG GGG CGGG ACT ATGG TTG CTG ACT AATT GAG ATG CATG CTT TGCA TAC TTCT GCCT GC
 130 140 150 160 170 180
 TGG GAG CCT GGG ACT TTG CAC AC TG GT TG CT GACT AA TT GAG ATG CA TG CT TT GCA T
 190 200 210 220 230 240
 ACT TCT GCCT GCT GGG GAGC CT GGG ACT TT TCC AC ACC CT AACT GAC ACA CATT CCAC AG
 250 260 270 280 290 300
 AATTA ATT CC CCT AG TT ATT AAT AG TAAT C AATT AC GGGG TC ATT AG TT TC AT AG CC CATA
 310 320 330 340 350 360
 TAT GGAG TT C CGC GTT ACAT AACT TAC GGT AA AT GGCC CG CCT GG CT GAC CG CCC AAC GA
 370 380 390 400 410 420
 CCC CGCC CA TTG AC GT CAA TAAT GAC GT A TG TT CC CATA GT AAC GCAA TAGG GACT TT
 430 440 450 460 470 480
 CATT GAC GT CA AT GGG TGG AG TAT TTAC G TGAA ACT G CC ACT TGG CAG TAC AT CA AGT
 490 500 510 520 530 540
 GT AT CAT AT G CCA AGT AC GC CCC CT ATT GA CGT CA AT GAC G GT AA AT GG GC CC CG CT GG CA
 550 560 570 580 590 600
 TT AT GCCC AG TAC AT GAC CT TAT GGG ACT T TC CT ACT TT GG CAG TAC AT CT AC GT ATT AGT
 610 620 630 640 650 660
 CAT CG CT ATT ACC AT GGT GA TG CG GTT TTG GC AGT AC AT C AAT GGG CGT G GAT AG CG GTT
 670 680 690 700 710 720
 TG ACT CAC GGG GAT TT CCAA GT CT CCAC CC CATT GAC GT C AAT GGG AG TT TG TT TT GA AG
 730 740 750 760 770 780
 GT T TAA AC AG CT TG GCG G CCAG CT TT A TTA AC GT GT TT AC GT CG AG TCA ATT GT AC
 790 800 810 820 830 840
 ACT A AC GACA GT GAT GAA AG AA AT AC AAAA GCG CATA ATA TTT TGA AC GA CG TCG AAC CT
 850 860 870 880 890 900
 TT ATT AC AAAA ACA AA AC ACA AAC GA AT AC GACA AG CT A GATT GCT GCT ACA AG AT TT G
 910 920 930 940 950 960
 GCA AGT TTG TGG CGT TGAG CG AAA ATCCA TT AGA TAG TC CAG CC AT CGG TT CGG AAAA
 970 980 990 1000 1010 1020
 CA ACC CT TG TT GAA ACT AA TCG AA AC CT TA TTT AC AA AT CT ATT GAG GA TTT A AT ATT
 1030 1040 1050 1060 1070 1080
 AA ATT CAG AT A TAA AGA AC GC TG AAA AT CA TT GAT TT CG CT CT A AC AT A CC ACC CT AAA
 1090 1100 1110 1120 1130 1140
 GATT ATTA AT TTA AT GA ATT ATT AAA AT AC AT CAG CA ACT AT AT ATT GAT AG AC AT TT CC
 1150 1160 1170 1180 1190 1200
 AG TTT GTG AT ATT AG TT GT GCG TCT CATT ACA AT GG CT G TT AT TT TT AA CA AC AA AC AA
 1210 1220 1230 1240 1250 1260
 CT GCT CGC AG ACA AT AGT AT AG AAA AGG GA GGT GAA CT GT TT TT GT TT AA CG GTT CG TAC
 1270 1280 1290 1300 1310 1320
 AAC AT TT GG AA AG TT AT GT TA AT CC GG TG CT GCA AAAA AT GG GT TA AT TG AACT AGA A

DNASIS
Mandy 5E8N-SHL

1330	1340	1350	1360	1370	1380
GAAGCTGCGT ACTATGCCGG CAACATATTG TACAAAACCG ACGATCCAA ATTCAATTGAT					
1390	1400	1410	1420	1430	1440
TATATAAATT TAATAATTAA AGCAACACAC TCCGAAGAAC TACAGAAAA TAGCACTGTT					
1450	1460	1470	1480	1490	1500
GTAAATTACA GAAAAACTAT GCGCAGCGGT ACTATACACC CCATTAAGAA AGACATATAT					
1510	1520	1530	1540	1550	1560
ATTTATGACA ACAAAAAATT TACTCTATAC GATAGATACA TATATGGATA CGATAATAAC					
1570	1580	1590	1600	1610	1620
TATGTTAATT TTTATGAGGA GAAAAATGAA AAAGAGAAGG AATACGAAGA AGAAGACGAC					
1630	1640	1650	1660	1670	1680
AAGGCCTCTA GTTTATGTGA AAATAAAATT ATATTGTCGC AAATTAAC TGATCATTT					
1690	1700	1710	1720	1730	1740
GAAAATGATT TAAATATTA CCTCAGCGAT TATAACTACG CGTTTCAAT TATAGATAAT					
1750	1760	1770	1780	1790	1800
ACTACAAATG TTCTTGTGC GTTTGGTTTG TATCGTTAAT AAAAACAAA TTTGACATT					
1810	1820	1830	1840	1850	1860
ATAATTGTTT TATTATCAA TAATTACAAA TAGGATTGAG ACCCTTGAG TTGCCAGCAA					
1870	1880	1890	1900	1910	1920
ACGGACAGAG CTTGTCGAGG AGAGTTGTTG ATTCAATTGTT TGCCCTCCCTG CTGCGGTTT					
1930	1940	1950	1960	1970	1980
TCACCGAAGT TCATGCCAGT CCAGCGTTT TGCAAGCAGAA AAGCCGCCGA CTTCGGTTT					
1990	2000	2010	2020	2030	2040
CGGTCGCGAG TGAAGATCCC TTTCTGTTA CCGCCAACGC GCAATATGCC TTGCGAGGTC					
2050	2060	2070	2080	2090	2100
GCAAAATCGG CGAAATTCCA TACCTGTTCA CCGACGACGG CGCTGACGCG ATCAAAGACG					
2110	2120	2130	2140	2150	2160
CGGTGATACA TATCCAGCCA TGCACACTGA TACTCTTCAC TCCACATGTC GGTGTACATT					
2170	2180	2190	2200	2210	2220
GAGTGCAGCC CGGCTAACGT ATCCACGCCG TATTCGGTGA TGATAATCGG CTGATGCAGT					
2230	2240	2250	2260	2270	2280
TTCTCCTGCC AGGCCAGAAG TTCTTTTCC AGTACCTTCT CTGCCGTTTC CAAATGCCG					
2290	2300	2310	2320	2330	2340
CTTGGACAT ACCATCCGTA ATAACGGTTC AGGCACAGCA CATCAAAGAG ATCGCTGATG					
2350	2360	2370	2380	2390	2400
GTATCGGTGT GAGCGTCGCA GAACATTACA TTGACCGAGG TGATCGGACG CGTCGGGTCG					
2410	2420	2430	2440	2450	2460
AGTTTACGCG TTGCTCCGC CAGTGGCGCG AAATATTCCC GTGCACCTTG CGGACGGGTA					
2470	2480	2490	2500	2510	2520
TCCGGTTCGT TGGCAATACT CCACATCACC ACGCTGGGT GGTTTTGTC ACGCGCTATC					
2530	2540	2550	2560	2570	2580
AGCTTTAA TCGCCTGTAA GTGCGCTTGC TGAGTTCCC CGTTGACTGC CTCTCGCTG					
2590	2600	2610	2620	2630	2640

DNASIS
Mandy + 5E8N-SHL

TACAGTTCTT TCGGCTTGT GCCCCGTTCG AAACCAATGC CTAAAGAGAG GTTAAAGGCCG
 2650 2660 2670 2680 2690 2700
 ACAGCAGCAG TTTCATCAAT CACCACGATG CCATGTTCAT CTGCCAGTC GAGCATCTCT
 2710 2720 2730 2740 2750 2760
 TCAGCGTAAG GGTAATGCGA GGTACGGTAG GAGTTGGCCC CAATCCAGTC CATTAATGCG
 2770 2780 2790 2800 2810 2820
 TGGTCGTGCA CCATCAGCAC GTTATCGAAT CCTTGCCAC GCAAGTCCGC ATCTTCATGA
 2830 2840 2850 2860 2870 2880
 CGACCAAAGC CAGTAAAGTA GAACGGTTTG TGGTTAATCA GGAACTGTTG GCCCTTCACT
 2890 2900 2910 2920 2930 2940
 GCCACTGACC GGATGCCGAC GCGAAGCGGG TAGATATCAC ACTCTGTCTG GCTTTGGCT
 2950 2960 2970 2980 2990 3000
 TGACGCACA GTTCATAGAG ATAACCTTCA CCCGGTTGCC AGAGGTGCCG ATTCAACACT
 3010 3020 3030 3040 3050 3060
 TGCAAAGTCC CGCTAGTGCC TTGTCCAGTT GCAACCACCT GTTGATCCGC ATCACCGAGT
 3070 3080 3090 3100 3110 3120
 TCAACGCTGA CATCACCAATT GGCCACCACCC TGCCAGTCAA CAGACGCGTG GTTACAGTCT
 3130 3140 3150 3160 3170 3180
 TCGCGACAT GCGTCACCAAC GGTGATATCG TCCACCCAGG TGTTGGCGT GGTGTAGAGC
 3190 3200 3210 3220 3230 3240
 ATTACGCTGC GATGGATTCC GGCATAGTTA AAGAAATCAT GGAAGTAAGA CTGCTTTTC
 3250 3260 3270 3280 3290 3300
 TTGCCGTTTT CGTCGGTAAT CACCATTCCC GGCGGGATAG TCTGCCAGTT CAGTTCGTTG
 3310 3320 3330 3340 3350 3360
 TCACACAAA CGGTGATACC CCTCGACGGA TTAAAGACTT CAAGCGGTCA ACTATGAAGA
 3370 3380 3390 3400 3410 3420
 AGTGTTCGTC TTCTGCCCAG TAAGCTATGT CTCCAGAATG TAGCCATCCA TCCTTGTCAA
 3430 3440 3450 3460 3470 3480
 TCAAGGC GTT GGTGCTTCC GGATTGTTTA CATAACCGGA CATAATCATA GGTCCCTCTGA
 3490 3500 3510 3520 3530 3540
 CACATAATTG GCCTCTCTGA TTAACGCCA GCGTTTTCCC GGTATCCAGA TCCACAAACCT
 3550 3560 3570 3580 3590 3600
 TCGCTTCAAA AAATGGAACA ACTTTACCGA CCGCGCCCGG TTTATCATCC CCCTCGGGTG
 3610 3620 3630 3640 3650 3660
 TAATCAGAAT AGCTGATGTA GTCTCAGTGA GCCCATATCC TTGTCGTATC CCTGGAAGAT
 3670 3680 3690 3700 3710 3720
 GGAAGCGTTT TGCAACCGCT TCCCCGACTT CTTTCGAAAG AGGTGCGCCC CCAGAAGCAA
 3730 3740 3750 3760 3770 3780
 TTTCGTGTAAT TAGAGATAAA TCGTATTGTG CAATCAGAGT GCTTTGGCG AAGAATGAAA
 3790 3800 3810 3820 3830 3840
 ATAGGGTTGG TACTAGCAAC GCACCTTGAA TTTTGTAAATC CTGAAGGGAT CGTAAAAAACAA
 3850 3860 3870 3880 3890 3900
 GCTCTTCTTC AAATCTATAC ATTAAGACGA CTCGAAATCC ACATATCAAAT TATCCGAGTG

DNASIS
Mandy + S8N-SHL

3910	3920	3930	3940	3950	3960
TAGTAAACAT TCCAAAACCG TGATGGAATG GAACAACACT TAAAATCGCA GTATCCGGAA					
3970	3980	3990	4000	4010	4020
TGATTTGATT GCCAAAAATA GGATCTCTGG CATGCGAGAA TCTGACGCAG GCAGTTCTAT					
4030	4040	4050	4060	4070	4080
GCGGAAGGGC CACACCCCTTA GGTAAACCCAG TAGATCCAGA GGAATTGTTT TGTACAGATC					
4090	4100	4110	4120	4130	4140
AAAGGACTCT GGTACAAAAT CGTATTCTATT AAAACCGGGGA GGTAGATGAG ATGTGACGAA					
4150	4160	4170	4180	4190	4200
CGTGTACATC GACTGAAATC CCTGGTAATC CGTTTGTAGAA TCCATGATAA TAATTTCTG					
4210	4220	4230	4240	4250	4260
GATTATTGGT AATTTTTTTT GCACGTTCAA AATTTTTTGC AACCCCTTTT TGAAACAAA					
4270	4280	4290	4300	4310	4320
.CTACGGTA GGCTGCGAAA TGTTCATACT GTTGAGCAAT TCACGTTCAT TATAATGTC					
4330	4340	4350	4360	4370	4380
GTTCGCGGGC GCAACTGCAA CTCCGATAAAA TAACCGCGCC AACACCGGCA TAAAGAATTG					
4390	4400	4410	4420	4430	4440
AAGAGAGTTT TCACTGCATA CGACGATTCT GTGATTTGTA TTCAGCCCCAT ATCGTTTCAT					
4450	4460	4470	4480	4490	4500
AGCTTCTGCC AACCGAACGG ACATTTGAA GTATTCGCG TACAGCCCCG CCGTTTAAAC					
4510	4520	4530	4540	4550	4560
GGCCGGGCTT CAATACCCCTG ATTGACTGGA ACAGCTGTAG CCCTGAACAG CAGCGTGC					
4570	4580	4590	4600	4610	4620
TGCTGACCGG TCCGGCGATT TCCGCCCTTG ACAGTATTAC CCGGACGGTC AGCGATATT					
4630	4640	4650	4660	4670	4680
.GATAATGT AAAAACGCGC GGTGACGATG CCCTGCGTGA ATACAGCGCT AAATTTGATA					
4690	4700	4710	4720	4730	4740
AAACAGAAGT GACAGCGCTA CGCGTCACCC CTGAAGAGAT CGCCGCCGCC GGCGCGCGTC					
4750	4760	4770	4780	4790	4800
TGAGCGACGA ATTAAAACAG GCGATGACCG CTGCCGTCAA AAATATTGAA ACGTTCCATT					
4810	4820	4830	4840	4850	4860
CCGCGCAGAC GCTACCGCCT GTAGATGTGG AAACCCAGCC AGGCCTGCGT TGCCAGCAGG					
4870	4880	4890	4900	4910	4920
TTACCGCGTCC CGTCTCGTCT GTCGGTCTGT ATATTCGGG CGGCTCGGCT CCGCTCTCT					
4930	4940	4950	4960	4970	4980
CAACGGTGCT GATGCTGGCG ACGCCGGCGC GCATTGCGGG ATGCCAGAAG GTGGTTCTGT					
4990	5000	5010	5020	5030	5040
GCTCGCCGCC GCCCATCGCT GATGAAATCC TCTATCGGC GCAACTGTGT GGCCTGCGAGG					
5050	5060	5070	5080	5090	5100
AAATCTTAA CGTCGGCGGC GCGCAGGCGA TTGCCGCTCT GGCCCTCGGC AGCGAGTCCG					
5110	5120	5130	5140	5150	5160
TACCGAAAGT GGATAAAATT TTTGGCCCCG GCAACGCCCT TGTAACCAGA GCCAACGTC					
5170	5180	5190	5200	5210	5220
AGGTCAAGCCA GCGTCTCGAC GCGCGGGCTA TCGATATGCC AGCCGGGCCG TCTGAAGTAC					

DNASI
Mandy 5E8N-SHL

5230	5240	5250	5260	5270	5280
TGGTGATCGC	AGACAGCGGC	GCAACACCGG	ATTCGTGCGC	TTCTGACCTG	CTCTCCAGG
5290	5300	5310	5320	5330	5340
CTGAGCACGG	CCCGGATTCC	CAGGTGATCC	TGCTGACGCC	TGATGCTGAC	ATTGCCCGCA
5350	5360	5370	5380	5390	5400
AGGTGGCGGA	GGCGGTAGAA	CGTCAACTGG	CGGAAC TGCC	GCGCGCGGAC	ACCGCCCCGGC
5410	5420	5430	5440	5450	5460
AGGCCCTGAG	CGCCAGTCGT	CTGATTGTGA	CCAAAGATT	AGCGCAGTGC	GTCGCCATCT
5470	5480	5490	5500	5510	5520
CTAATCAGTA	TGGGCCGGAA	CACTTAATCA	TCCAGACGCC	CAATGCGCGC	GATTTGGTGG
5530	5540	5550	5560	5570	5580
ATGCGATTAC	CAGCGCAGGC	TCGGTATTTC	TCGGCGACTG	GTCGCCGGAA	TCCGCCGGTG
5590	5600	5610	5620	5630	5640
ATTACGCTTC	CGGAACCAAC	CATGTTTAC	CGACCTATGG	CTATACTGCT	ACCTGTTCCA
5650	5660	5670	5680	5690	5700
GCCTTGGGTT	AGCGGATTC	CAGAAACGGA	TGACCGTTCA	GGAACTGTG	AAAGCGGGCT
5710	5720	5730	5740	5750	5760
TTTCCGCTCT	GGCATCAACC	ATTGAAACAT	TGGCGCGGC	AGAACGTCTG	ACCGCCCATA
5770	5780	5790	5800	5810	5820
AAAATGCCGT	GACCCCTGCGC	GTAAACGCC	TCAAGGAGCA	AGCATGAGCA	CTGAAAACAC
5830	5840	5850	5860	5870	5880
TCTCAGCGTC	GCTGACTTAG	CCCGTGAAAA	TGTCCGCAAC	CTGGAGATCC	AGACATGGAT
5890	5900	5910	5920	5930	5940
AAGATACATT	GATGAGTTT	GACAAACAC	AACTAGAATG	CAGTGAAAAA	AATGCTTAT
5950	5960	5970	5980	5990	6000
TTGTGAAATT	TGTGATGCTA	TTGCTTTATT	TGTAACCATT	ATAAGCTGCA	ATAAACAAAGT
6010	6020	6030	6040	6050	6060
TAACAACAAC	AATTGCATTC	ATTTTATGTT	TCAGGTTCA	GGGGAGGTGT	GGGAGGTTTT
6070	6080	6090	6100	6110	6120
TTAAAGCAAG	TAAAACCTCT	ACAAATGTGG	TATGGCTGAT	TATGATCTCT	AGGGCCGGCC
6130	6140	6150	6160	6170	6180
CTCGACGGCG	CGTCTAGAGC	AGTGTGGTT	TCAAGAGGAA	GCAAAAGCC	TCTCCACCCA
6190	6200	6210	6220	6230	6240
GGCCTGGAAT	GTTTCCACCC	AATGTCGAGC	AGTGTGGTT	TGCAAGAGGA	AGCAAAAGC
6250	6260	6270	6280	6290	6300
CTCTCCACCC	AGGCCTGGAA	TGTTTCCACC	CAATGTCGAG	CAAACCCCGC	CCAGCGTCTT
6310	6320	6330	6340	6350	6360
GTCATTGGCG	AATTGGAACA	CGCATATGCA	GTCGGGGCGG	CGCGGTCCCA	GGTCCACTTC
6370	6380	6390	6400	6410	6420
GCATATTAAG	GTGGCGCGTG	TGGCCTCGAA	CACCGAGCGA	CCCTGCA GCC	AATATGGGAT
6430	6440	6450	6460	6470	6480
CGGCCATTGA	ACAAGATGGA	TTGCACGCCAG	GTTCTCCGGC	CGCTTGGGTG	GAGAGGCTAT
6490	6500	6510	6520	6530	6540

DNASIS
Mandy - E8N-SHL

TCGGCTATGA CTGGGCACAA CAGACAATCG GCTGCTCTGA TGCCGCCGTG TTCCGGCTGT
 6550 6560 6570 6580 6590 6600
 CAGCGCAGGG GCGCCCGGTT CTCCCCGTCA AGACCGACCT GTCCGGTGCC CTGAATGAAC
 6610 6620 6630 6640 6650 6660
 TGCAGGTAAG TGCAGGCCGTC GATGGCCGAG GCGGCCCTCGG CCTCTGCATA AATAAAAAAA
 6670 6680 6690 6700 6710 6720
 ATTAGTCAGC CATGCATGGG GCGGAGAATG GGCAGGACTG GCGGAGTTA GGGCGGGAT
 6730 6740 6750 6760 6770 6780
 GGGCGGAGTT AGGGGCGGGA CTATGGTTGC TGACTAATTG AGATGCATGC TTTGCATACT
 6790 6800 6810 6820 6830 6840
 TCTGCCTGCT GGGGAGCCTG GGGACTTTC ACACCTGGTT GCTGACTAAT TGAGATGCAT
 6850 6860 6870 6880 6890 6900
 CCTTTGCATA CTTCTGCCTG CTGGGGAGCC TGGGGACTTT CCACACCCCTA ACTGACACAC
 6910 6920 6930 6940 6950 6960
 ATTCCACAGA ATTAATTCCC CTAGTTATTA ATAGTAATCA ATTACGGGGT CATTAGTTCA
 6970 6980 6990 7000 7010 7020
 TAGCCCATAT ATGGAGTTCC GCGTTACATA ACTTACGGTA AATGGCCCCGC CTGGCTGACC
 7030 7040 7050 7060 7070 7080
 GCCCAACGAC CCCCGCCCAT TGACGTCAAT AATGACGTAT GTTCCCATAG TAACGCCAAT
 7090 7100 7110 7120 7130 7140
 AGGGACTTTC CATTGACGTC AATGGGTGGA GTATTTACGG TAAACTGCC ACTTGGCAGT
 7150 7160 7170 7180 7190 7200
 ACATCAAGTG TATCATAATGC CAAGTACGCC CCCTATTGAC GTCAATGACG GTAAATGGCC
 7210 7220 7230 7240 7250 7260
 CCTCTGGCAT TATGCCCACT ACATGACCTT ATGGGACTTT CCTACTTGGC AGTACATCTA
 7270 7280 7290 7300 7310 7320
 CGTATTAGTC ATCGCTATTA CCATGGTGAT GCGGTTTTGG CAGTACATCA ATGGCGTGG
 7330 7340 7350 7360 7370 7380
 ATAGCGGTTT GACTCACGGG GATTTCCAAG TCTCCACCCC ATTGACGTCA ATGGGAGTTT
 7390 7400 7410 7420 7430 7440
 GTTTTGGCAC CAAAATCAAC GGGACTTTCC AAAATGTCGT ACAACTCCG CCCCATGAC
 7450 7460 7470 7480 7490 7500
 GCAAATGGGC GGTAGGCCTG TACGGTGGGA GGTCTATATA AGCAGAGCTG GGTACGTGAA
 7510 7520 7530 7540 7550 7560
 CCGTCAGATC GCCTGGAGAC GCCATCACAG ATCTCTCACC ATGGACATGA GGGTCCCCGC
 7570 7580 7590 7600 7610 7620
 TCAGCTCCTG GGGCTCCTTC TGCTCTGGCT CCCAGGTGCC AGATGTGACA TCCAGATGAC
 7630 7640 7650 7660 7670 7680
 CCAGTCTCCA TCTTCCCTGT CTGCATCTGT AGGGGACAGA GTCACCATCA CTTGCAGGGC
 7690 7700 7710 7720 7730 7740
 AAGTCAGGAC ATTAGGTATT ATTTAAATTG GTATCAGCAG AAACCAAGGAA AAGCTCCTAA
 7750 7760 7770 7780 7790 7800
 GCTCCCTGATC TATGTTGCAT CCAGTTGCA AAGTGGGGTC CCATCAAGGT TCAGCGGCAG

DNASIS
Mandy EBN-SHL

7810	7820	7830	7840	7850	7860
TGGATCTGGG ACAGAGTTCA CTCTCACCGT CAGCAGCTG CAGCCTGAAG ATTTTCGAC					
7870	7880	7890	7900	7910	7920
TTATTACTGT CTACAGGTTT ATAGTACCCC TCGGACGTT GCCTAAAGGG ACAAAGGTGGA					
7930	7940	7950	7960	7970	7980
AATCAAACGT ACGGTGGCTG CACCATCTGT CTTCATCTTC CCGCCATCTG ATGAGCAGTT					
7990	8000	8010	8020	8030	8040
GAAATCTGGA ACTGCCTCTG TTGTGTGCCT GCTGAATAAC TTCTATCCCA GAGAGGCCAA					
8050	8060	8070	8080	8090	8100
AGTACAGTGG AAGGTGGATA ACGCCCTCCA ATCGGGTAAC TCCCAGGAGA GTGTACAGA					
8110	8120	8130	8140	8150	8160
GCAGGACAGC AAGGACAGCA CCTACAGCCT CAGCAGCACC CTGACGCTGA GCAAAGCAGA					
8170	8180	8190	8200	8210	8220
TACGAGAAA CACAAAGTCT ACGCCTGCAG AGTCACCCAT CAGGGCCTGA GCTCGCCCGT					
8230	8240	8250	8260	8270	8280
CACAAAGAGC TTCAACAGGG GAGAGTGTG AATTCAAGATC CGTTAACGGT TACCAACTAC					
8290	8300	8310	8320	8330	8340
CTAGACTGGA TTCGTGACAA CATGCGGCCG TGATATCTAC GTATGATCAG CCTCGACTGT					
8350	8360	8370	8380	8390	8400
GCCTTCTAGT TGCCAGCCAT CTGTTGTTG CCCCTCCCCC GTGCCCTCT TGACCCCTGGA					
8410	8420	8430	8440	8450	8460
AGGTGCCACT CCCACTGTCC TTTCTTAATA AAATGAGGAA ATTGCATCGC ATTGTCTGAG					
8470	8480	8490	8500	8510	8520
TAGGTGTCAAT TCTATTCTGG GGGGTGGGGT GGGGCAGGAC AGCAAGGGGG AGGATTGGGA					
8530	8540	8550	8560	8570	8580
ACAATAGC AGGCATGCTG GGGATGCGGT GGGCTCTATG GCTTCTGAGG CGGAAAGAAC					
8590	8600	8610	8620	8630	8640
CAGCTGGGAC TAGTCGAAT TGGGGGGAGT TAGGGGCGGG ATGGGGCGGAG TTAGGGGCGG					
8650	8660	8670	8680	8690	8700
GACTATGGTT GCTGACTAAT TGAGATGCAT GCTTGCATA CTTCTGCCTG CTGGGGAGCC					
8710	8720	8730	8740	8750	8760
TGGGGACTTT CCACACCTGG TTGCTGACTA ATTGAGATGC ATGCTTTGCA TACTTCTGCC					
8770	8780	8790	8800	8810	8820
TGCTGGGGAG CCTGGGGACT TTCCACACCC TAACTGACAC ACATTCCACA GAATTAATT					
8830	8840	8850	8860	8870	8880
CCCTAGTTAT TAATAGTAAT CAATTACGGG GTCAATTAGTT CATAGCCCAT ATATGGAGTT					
8890	8900	8910	8920	8930	8940
CCGCCTTACA TAACTTACGG TAAATGGCCC GCCTGGCTGA CCGCCCAACG ACCCCCCGCC					
8950	8960	8970	8980	8990	9000
ATTGACGTCA ATAATGACGT ATGTTCCCAT AGTAACGCCA ATAGGGACTT TCCATTGACG					
9010	9020	9030	9040	9050	9060
TCAATGGGTG GAGTATTTCAC GGTAAACTGC CCACTGGCA GTACATCAAG TGTATCATAT					
9070	9080	9090	9100	9110	9120
GCCAAGTACG CCCCCATTG ACGTCAATGA CGGTAAATGG CCGGCTGGC ATTATGCCCA					

DNASIS
Mandy 5E8N-SHL

9130	9140	9150	9160	9170	9180
GTACATGACC	TTATGGGACT	TTCCTACTTG	GCAGTACATC	TACGTATTAG	TCATCGCTGT
9190	9200	9210	9220	9230	9240
TACCATGGTG	ATGCGGTTTT	GGCAGTACAT	CAATGGCGT	GGATAGCGGT	TTGACTCACG
9250	9260	9270	9280	9290	9300
GGGATTTCCA	AGTCTCCACC	CCATTGACGT	CAATGGGAGT	TTGTTTGGC	ACCAAAATCA
9310	9320	9330	9340	9350	9360
ACGGGACTTT	CCAAAATGTC	GTAACAACTC	CGCCCCATTG	ACGCAAATGG	GCGTAGGCG
9370	9380	9390	9400	9410	9420
TGTACGGTGG	GAGGTCTATA	TAAGCAGAGC	TGGGTACGTG	AACCGTCAGA	TCGCCCTGGAG
9430	9440	9450	9460	9470	9480
ACGCCGTCGA	CATGGGTTGG	AGCCTCATCT	TGCTCTTCCT	TGTGCTGTGTT	GCTACGCGTG
9490	9500	9510	9520	9530	9540
. CCTGTCCGA	GGTGCAGCTG	GTGGAGCTG	GGGGCGGCTT	GGCAAAGCCT	GGGGGGTCCC
9550	9560	9570	9580	9590	9600
TGAGACTCTC	CTGCGCAGCC	TCCGGGTTCA	GGTTCACCTT	CAATAACTAC	TACATGGACT
9610	9620	9630	9640	9650	9660
GGGTCCGCCA	GGCTCCAGGG	CAGGGGCTGG	AGTGGGTCTC	ACGTATTAGT	AGTAGTGGTG
9670	9680	9690	9700	9710	9720
ATCCCACATG	GTACGCAGAC	TCCGTGAAGG	GCAGATTAC	CATCTCCAGA	GAGAACGCCA
9730	9740	9750	9760	9770	9780
AGAACACACT	GTTTCTCAA	ATGAACAGCC	TGAGAGCTGA	GGACACGGCT	GTCTATTACT
9790	9800	9810	9820	9830	9840
GTGCGAGCTT	GAECTACAGGG	TCTGACTCCT	GGGGCCAGGG	AGTCCTGGTC	ACCGTCTCCT
9850	9860	9870	9880	9890	9900
LAGCTAGCAC	CAAGGGCCA	TCGGTCTTCC	CCCTGGCACC	CTCCCTCCAAG	AGCACCTCTG
9910	9920	9930	9940	9950	9960
GGGGCACAGC	GGCCCTGGGC	TGCCTGGTCA	AGGACTACTT	CCCCGAACCG	GTGACGGTGT
9970	9980	9990	10000	10010	10020
CGTGGAACTC	AGGCGCCCTG	ACCAGCAGCG	TGCACACCTT	CCCGGCTGTC	CTACAGTCCT
10030	10040	10050	10060	10070	10080
CAGGACTCTA	CTCCCTCAGC	AGCGTGGTGA	CCGTGCCCTC	CAGCAGCTTG	GGCACCCAGA
10090	10100	10110	10120	10130	10140
CCTACATCTG	CAACGTGAAT	CACAAGCCA	GCAACACAA	GGTGGACAAG	AAAGTTGAGC
10150	10160	10170	10180	10190	10200
CCAAATCTG	TGACAAAACT	CACACATGCC	CACCGTGCCTC	AGCACCTGAA	CTCCTGGGGG
10210	10220	10230	10240	10250	10260
GACCGTCAGT	CTTCCTCTTC	CCCCAAAAC	CCAAGGACAC	CCTCATGATC	TCCCGGACCC
10270	10280	10290	10300	10310	10320
CTGAGGTAC	ATGCGTGGTG	GTGGACGTGA	GCCACGAAGA	CCCTGAGGTC	AAGTTCAACT
10330	10340	10350	10360	10370	10380
GGTACGTGGA	CGGCGTGGAG	GTGCATAATG	CCAAGACAA	GCCGCGGGAG	GAGCAGTACA
10390	10400	10410	10420	10430	10440

DNASIS
Mandy SE8N-SHL

ACAGCACGTA CGGTGTGGTC AGCGTCTCA CGTCCTGCA CCAGGACTGG CTGAATGGCA

10450 10460 10470 10480 10490 10500
AGGAGTACAA GTGCAAGGTC TCCAACAAAG CCCTCCCAGC CCCCATCGAG AAAACCATCT

10510 10520 10530 10540 10550 10560
CCAAAGCCAA AGGGCAGCCC CGAGAACAC AGGTGTACAC CCTGCCCCA TCCCGGGATG

10570 10580 10590 10600 10610 10620
AGCTGACCAA GAACCAGGTC AGCCTGACCT GCCTGGTCAA AGGCTTCTAT CCCAGCGACA

10630 10640 10650 10660 10670 10680
TCGCCGTGGA GTGGGAGAGC AATGGGCAGC CGGAGAACAA CTACAAGACC ACGGCTCCCG

10690 10700 10710 10720 10730 10740
TGCTGGACTC CGACGGCTCC TTCTTCCTCT ACAGCAAGCT CACCGTGGAC AAGAGCAGGT

10750 10760 10770 10780 10790 10800
GCAGCAGGG GAACGTCTTC TCATGCTCCG TGATGCATGA GGCTCTGCAC AACCACTACA

10810 10820 10830 10840 10850 10860
CGCAGAAAGAG CCTCTCCCTG TCTCCGGGTA AATGAGGATC CGTTAACGGT TACCAACTAC

10870 10880 10890 10900 10910 10920
CTAGACTGGA TTCGTGACAA CATGCGGGCG TGATATCTAC GTATGATCAG CCTCGACTGT

10930 10940 10950 10960 10970 10980
GCCTTCTAGT TGCCAGCCAT CTGTTGTTTG CCCCTCCCCC GTGCCTTCCCT TGACCCCTGGA

10990 11000 11010 11020 11030 11040
AGGTGCCACT CCCACTGTCC TTTCTTAATA AAATGAGGAA ATTGCATCGC ATTGTCTGAG

11050 11060 11070 11080 11090 11100
TAGGTGTCA TCTATTCTGG GGGGTGGGGT GGGGCAGGAC AGCAAGGGGG AGGATTGGGA

11110 11120 11130 11140 11150 11160
ACAATAGC AGGCATGCTG GGGATGCCGT GGGCTCTATG GCTTCTGAGG CGGAAAGAAC

11170 11180 11190 11200 11210 11220
CAGCTGGGGC TCGACAGCAA CGCTAGGTG AGGCCGCTAC TAACTCTCTC CTCCCTCCTT

11230 11240 11250 11260 11270 11280
TTCTTGCA GACGAGGCAG CGCGGTATC GTGGCTGGCC ACGACGGGCG TTCTTGCGC

11290 11300 11310 11320 11330 11340
AGCTGTGCTC GACGTGTCA CTGAAGCGGG AAGGGACTGG CTGCTATTGG GCGAAGTGCC

11350 11360 11370 11380 11390 11400
GGGGCAGGAT CTCTGTCA CTCACCTTGC TCCTGCCAG AAAGTATCCA TCATGGCTGA

11410 11420 11430 11440 11450 11460
TGCAATGCGG CGGCTGCATA CGCTTGATCC GGCTACCTGC CCATTCGACC ACCAAGCGAA

11470 11480 11490 11500 11510 11520
ACATCGCATC GAGCGAGCAC GTACTCGGAT GGAAGCCGGT CTTGTCGATC AGGATGATCT

11530 11540 11550 11560 11570 11580
GGACGAAGAG CATCAGGGGC TCGCGCCAGC CGAACTGTTC GCCAGGTAAG TGAGCTCCAA

11590 11600 11610 11620 11630 11640
TTCAAGCTCT CGAGCTAGGG CGGCCAGCTA GTAGCTTGC TTCTCAATT CTTATTGCA

11650 11660 11670 11680 11690 11700
TAATGAGAAA AAAAGGAAAA TTAATTTAA CACCAATTCA GTAGTTGATT GAGCAAATGC

DNASIS
Mandy SE8N-SHL

11710	11720	11730	11740	11750	11760
GTTGCCAAA AGGATGCTT AGAGACAGTG TTCTCTGCAC AGATAAGGAC AACATTATT					
11770	11780	11790	11800	11810	11820
CAGAGGGAGT ACCCAGAGCT GAGACTCCTA AGCCAGTGAG TGGCACAGCA TCCAGGGAGA					
11830	11840	11850	11860	11870	11880
AATATGCTTG TCATCACCGA AGCCTGATT CGTAGAGCCA CACCTGGTA AGGGCCAATC					
11890	11900	11910	11920	11930	11940
TGCTCACACA GGATAGAGAG GGCAGGAGCC AGGGCAGAGC ATATAAGGTG AGGTAGGATC					
11950	11960	11970	11980	11990	12000
AGTTGCTCCT CACATTTGCT TCTGACATAG TTGTGTTGGG AGCTTGGATA GCTTGGGGGG					
12010	12020	12030	12040	12050	12060
GGGACAGCTC AGGGCTGCGA TTTCGCCCA AACTTGACGG CAATCCTAGC GTGAAGGCTG					
12070	12080	12090	12100	12110	12120
AGGATTTT ATCCCCGCTG CCATCATGGT TCGACCATTG AACTGCATCG TCGCCGTGTC					
12130	12140	12150	12160	12170	12180
.CCAAAATATG GGGATTGGCA AGAACGGAGA CCTACCCCTGG CCTCCGCTCA GGAACGAGTT					
12190	12200	12210	12220	12230	12240
CAAGTACTTC CAAAGAATGA CCACAACCTC TTCAGTGGAA GGTAAACAGA ATCTGGTGAT					
12250	12260	12270	12280	12290	12300
TATGGGTAGG AAAACCTGGT TCTCCATTCC TGAGAAGAAT CGACCTTAA AGGACAGAAT					
12310	12320	12330	12340	12350	12360
TAATATAGTT CTCAGTAGAG AACTCAAAGA ACCACCAACGA GGAGCTCATT TTCTGCCAA					
12370	12380	12390	12400	12410	12420
AAGTTTGAT GATGCCCTAA CGTAGGGCGCG CCATTAAGAC TTATTGAACA ACCGGAATTG					
12430	12440	12450	12460	12470	12480
.CAAGTAAAG TAGACATGGT TTGGATAGTC GGAGGCAGTT CTGTTTACCA GGAAGCCATG					
12490	12500	12510	12520	12530	12540
AATCAACCAAG GCCACCTCAG ACTCTTTGTG ACAAGGATCA TGCAGGAATT TGAAAGTGAC					
12550	12560	12570	12580	12590	12600
ACGTTTTTCC CAGAAATTGA TTTGGGGAAA TATAAACTTC TCCCAGAATA CCCAGGCCTC					
12610	12620	12630	12640	12650	12660
CTCTCTGAGG TCCAGGAGGA AAAAGGCATC AAGTATAAGT TTGAAGTCTA CGAGAAGAAA					
12670	12680	12690	12700	12710	12720
GACTAACAGG AAGATGCTT CAAGTTCTCT GCTCCCCCTCC TAAAGCTATG CATTTTTATA					
12730	12740	12750	12760	12770	12780
AGACCATGGG ACTTTGCTG GCTTAGATC AGCTCGACT GTGCCTCTA GTGCCAGCC					
12790	12800	12810	12820	12830	12840
ATCTGTTGTT TGCCCTCCC CCGTGCCTTC CTTGACCTTG GAAGGTGCCA CTCCCACTGT					
12850	12860	12870	12880	12890	12900
CCTTTCTAA TAAAATGAGG AAATTGCATC GCATTGTCTG AGTAGGTGTC ATTCTATTCT					
12910	12920	12930	12940	12950	12960
GGGGGGTGGG GTGGGGCAGG ACAGCAAGGG GGAGGATTGG GAAGACAATA GCAGGCATGC					
12970	12980	12990	13000	13010	13020
TGGGGATGCG GTGGGCTCTA TGGCTTCTGA GGCGGAAAGA ACCAGCTGGG GCTCGAAGCG					

DNASIS
Mandy EBN-SHL

13030	13040	13050	13060	13070	13080
GCCGCCATT	TCGCTGGTGG	TCAGATGCGG	GATGGCGTGG	GACGCCGG	GGAGCGTCAC
13090	13100	13110	13120	13130	13140
ACTGAGGTTT	TCCGCCAGAC	GCCACTGCTG	CCAGGGCCTG	ATGTGCCCGG	CTTCTGACCA
13150	13160	13170	13180	13190	13200
TGCGGTCGCG	TTCGGTTGCA	CTACCGTAC	TGTGAGCCAG	AGTTGCCCGG	CGCTCTCCGG
13210	13220	13230	13240	13250	13260
CTGCGTAGT	TCAGGCAGTT	CAATCAACTG	TTTACCTTGT	GGAGCGACAT	CCAGAGGCAC
13270	13280	13290	13300	13310	13320
TTCACCGCTT	GCCAGCGGCT	TACCATCCAG	CGCCACCATC	CAGTGCAGGA	GCTCGTTATC
13330	13340	13350	13360	13370	13380
GCTATGACGG	AACAGGTATT	CGCTGGTCAC	TTCGATGGTT	TGCCCAGATA	AACGGAACGT
13390	13400	13410	13420	13430	13440
AAAACATGC	TGCTGGTGTT	TTGCTTCCGT	CAGCGCTGGA	TGCGGCGTGC	GGTCGGCAA
13450	13460	13470	13480	13490	13500
GACCAGACCG	TTCATACAGA	ACTGGCGATC	GTTCGGCGTA	TCGCCAAAT	CACCGCCGTA
13510	13520	13530	13540	13550	13560
AGCCGACAC	GGGTTGCCGT	TTTCATCATA	TTAACATCAGC	GACTGATCCA	CCCAGTCCC
13570	13580	13590	13600	13610	13620
GACGAAGCCG	CCCTGTAAAC	GGGGATACTG	ACGAAACGCC	TGCCAGTATT	TAGCGAAACC
13630	13640	13650	13660	13670	13680
GCCAAGACTG	TTACCCATCG	CGTGGGCGTA	TTCGCAAAGG	ATCAGCGGGC	GCCTCTCTCC
13690	13700	13710	13720	13730	13740
AGGTAGCGAA	AGCCATTTTT	TGATGGACCA	TTTCGGCACA	GCCGGGAAGG	GCTGGTCTTC
13750	13760	13770	13780	13790	13800
~TCCACCGCGC	GCGTACATCG	GGCAAATAAT	ATCGGTGGCC	GTGGTGTGCG	CTCCGCCGCC
13810	13820	13830	13840	13850	13860
TTCATACTGC	ACCGGGCGGG	AAGGATCGAC	AGATTTGATC	CAGCGATACA	GCGCGTCGTG
13870	13880	13890	13900	13910	13920
ATTAGCGCCG	TGGCCTGATT	CATTCCCCAG	CGACCAGATG	ATCACACTCG	GGTGATTACG
13930	13940	13950	13960	13970	13980
ATCGCGCTGC	ACCATTGCG	TTACCGTTC	GCTCATCGCC	GGTAGCCAGC	GCGGATCATC
13990	14000	14010	14020	14030	14040
GGTCAGACGA	TTCATTGGCA	CCATGCCGTG	GGTTTCAATA	TTGGCTTCAT	CCACACATA
14050	14060	14070	14080	14090	14100
CAGGCCGTAG	CGGTCGCACA	GCGTGTACCA	CAGCGGATGG	TTCGGATAAT	GCGAACAGCG
14110	14120	14130	14140	14150	14160
CACGGCGTTA	AAAGTTGTTCT	GCTTCATCAG	CAGGATATCC	TGCACCATCG	TCTGCTCATC
14170	14180	14190	14200	14210	14220
CATGACCTGA	CCATGCAGAG	GATGATGCTC	GTGACGGTTA	ACGCCCTGAA	TCAGCAACGG
14230	14240	14250	14260	14270	14280
CTTGCCGTTTC	AGCAGCAGCA	GACCATTTC	AATCCGCACC	TCGCGGAAAC	CGACATCGCA
14290	14300	14310	14320	14330	14340

DNASIS
Mandy E8N-SHL

GGCTTCTGCT TCAATCAGCG TGCCGTGGC GGTGTGCA GT TCAACCACCG CACGATAGAG

14350 14360 14370 14380 14390 14400
ATTGGGATT TCGGCCTCC ACAGTTTCGG GTTTGACG TTCAAGACGTA GTGTGACGCG

14410 14420 14430 14440 14450 14460
ATCGGCATAA CCACCA CGCT CATCGATAAT TTCACCGCCG AAAGGCGCGG TGCCGCTGGC

14470 14480 14490 14500 14510 14520
GACCTGCGTT TCACCCCTGCC ATAAAGAAC TGTTACCGT AGGTAGTCAC GCAACTCGCC

14530 14540 14550 14560 14570 14580
GCACATCTGA ACTTCAGCCT CCAGTACAGC GC GGCTGAAA TCATCATTAA AGCGAGTGGC

14590 14600 14610 14620 14630 14640
AACATGGAAA TCGCTGATTT GTGTAGTCGG TTTATGCAGC AACGAGACGT CACGGAAAAT

14650 14660 14670 14680 14690 14700
CGCTCATC CGCCACATAT CCTGATCTTC CAGATAACTG CCGTCACTCC AGCGCAGCAC

14710 14720 14730 14740 14750 14760
CATCACCGCG AGGC GGTTT CTCCGGCGCG TAAAAATGCG CTCAGGTCAA ATTCA GACGG

14770 14780 14790 14800 14810 14820
CAAACGACTG TCCTGGCCGT AACCGACCCA GCGCCCGTTG CACCA CAGAT GAAACGCCGA

14830 14840 14850 14860 14870 14880
GTTAACGCCA TCAAAATAA TTGCGTCTG GCCTTCTGT AGCCAGCTT CATCAACATT

14890 14900 14910 14920 14930 14940
AAATGTGAGC GAGTAACAAAC CCGTCGGATT CTCCGGGA ACAAACGGCG GATTGACCGT

14950 14960 14970 14980 14990 15000
AATGGGATAG GTCACGTTGG TGTAGATGGG CGCATCGTAA CCGTGCATCT GCCAGTTGA

15010 15020 15030 15040 15050 15060
GGACGACG ACAGTATCGG CCTCAGGAAG ATCGCACTCC AGCCAGCTT CCGGCACCGC

15070 15080 15090 15100 15110 15120
TTCTGGTGCC GGAAACCAGG CAAAGGCCA TTGCCATTG AGGCTGCGCA ACTGTTGGGA

15130 15140 15150 15160 15170 15180
AGGGCGATCG GTGCGGGCCT CTTGCTATT ACGCCAGCTG GCGAAAGGGG GATGTGCTGC

15190 15200 15210 15220 15230 15240
AAGGCGATTA AGTTGGGTAA CGCCAGGGTT TTCCCA GTCA CGACGTTGTA AAACGACTTA

15250 15260 15270 15280 15290 15300
ATCCGTGAG GGGCTGCCCTC GAAGCAGACG ACCTCCGTT GTGCAGCCAG CGGGCCCTGC

15310 15320 15330 15340 15350 15360
GCCGGTGCCC ACAATCGTGC GCGAACAAAC TAAACCAAGAA CAAATTATAC CGGGCGCAC

15370 15380 15390 15400 15410 15420
GCCGCCACCA CCTTCTCCCG TGCCTAACAT TCCAGCGCT CCACCAACAC CACCAACATC

15430 15440 15450 15460 15470 15480
GATGTCTGAA TTGCGGCCCG CTCCACCAAT GCGACGGAA CCTCAACCCG CTGCACCTTT

15490 15500 15510 15520 15530 15540
AGACGACAGA CAACAATTGT TGGAGCTAT TAGAACGAA AAAATCGCA CTCGTCTCAG

15550 15560 15570 15580 15590 15600
ACCGGTCAAA CCAAAACGG CGCCGAAAC CAGTACAATA GTTGAGGTGC CGACTGTGTT

DNASIS
Mandy E8N-SHL

15610 15620 15630 15640 15650 15660
 GCCTAAAGAG ACATTTGAGC CTAAACGCC GTCTGCATCA CCGCCACAC CTCCGCCTCC

 15670 15680 15690 15700 15710 15720
 GCCTCCGCCG CCAGCCCCGC CTGCGCTCC ACCGATGGTA GATTATCAT CAGCTCCACC

 15730 15740 15750 15760 15770 15780
 ACCGCCGCCA TTAGTAGATT TGCCGTCTGA AATGTTACCA CCGCCTGCAC CATCGCTTTC

 15790 15800 15810 15820 15830 15840
 TAACGTGTTG TCTGAATTAA AATCGGGCAC AGTTAGATTG AAACCCGCC AAAAACGCC

 15850 15860 15870 15880 15890 15900
 GCAATCAGAA ATAATTCAA AAAGCTAAC TACAAATTG ATCGCGGACG TGTAGCCGA

 15910 15920 15930 15940 15950 15960
 CACAATTAAT AGGCCTCGTG TGGCTATGGC AAAATCGTCT TCGGAAGCAA CTTCTAACGA

 15970 15980 15990 16000 16010 16020
 AGGGTTGG GACGACGACG ATAATCGGCC TAATAAGCT AACACGCCCG ATGTTAAATA

 16030 16040 16050 16060 16070 16080
 .TGTCCAAGCT ACTAGTGGTA CCGCTTGGCA GAACATATCC ATCGCGTCCG CCATCTCCAG

 16090 16100 16110 16120 16130 16140
 CAGCCGCACG CGGCGCATCT CGGGCAGCGT TGGGTCTGG CCACGGGTGC GCATGATCGT

 16150 16160 16170 16180 16190 16200
 GCTCCCTGTCG TTGAGGACCC GGCTAGGCTG GCGGGGTTGC CTTACTGGTT AGCAGAACGA

 16210 16220 16230 16240 16250 16260
 ATCACCGATA CGCGAGCGAA CGTGAAGCGA CTGCTGCTGC AAAACGTCG CGACCTGAGC

 16270 16280 16290 16300 16310 16320
 AACAAACATGA ATGGCTTCG GTTCCGTGT TTCGTAAAGT CTGGAAACGC GGAAGTCAGC

 16330 16340 16350 16360 16370 16380
 CCTGCACC ATTATGTTCC GGATCTGCAT CGCAGGATGC TGCTGGCTAC CCTGTGGAAC

 16390 16400 16410 16420 16430 16440
 ACCTACATCT GTATTAACGA AGCGCTGGCA TTGACCCCTGA GTGATTTTC TCTGGTCCCG

 16450 16460 16470 16480 16490 16500
 CCGCATCCAT ACCGCCAGTT GTTACCCCTC ACAACGTTCC AGTAACCGGG CATGTTCATC

 16510 16520 16530 16540 16550 16560
 ATCAGTAACC CGTATCGTGA GCATCCTCTC TCGTTTCATC GGTATCATTA CCCCCATGAA

 16570 16580 16590 16600 16610 16620
 CAGAAATCCC CCTTACACGG AGGCATCACT GACCAACAG GAAAAAACCG CCCTTAACAT

 16630 16640 16650 16660 16670 16680
 GGGCCGCTTT ATCAGAACGCC AGACATTAAC GCTTCTGGAG AACTCAACG AGCTGGACGC

 16690 16700 16710 16720 16730 16740
 GGATGAACAG GCAGACATCT GTGAATCGCT TCACGACAC GCTGATGAGC TTTACCGCAG

 16750 16760 16770 16780 16790 16800
 CTGCCTCGCG CGTTTCGGTG ATGACGGTGA AACCTCTGA CACATGCAGC TCCCGGAGAC

 16810 16820 16830 16840 16850 16860
 GGTACACAGCT TGTCTGTAAG CGGATGCCGG GAGCAGACAA GCCCGTCAGG GCGCGTCAGC

 16870 16880 16890 16900 16910 16920
 GGGTGTGGC GGGTGTGGG GCGCAGCCAT GACCCAGTCA CGTAGCGATA GCGGAGTGTAA

DNASIS
Mandy SE8N-SHL

16930 16940 16950 16960 16970 16980
 TACTGGCTTA ACTATGCGGC ATCAGAGCAG ATTGTACTGA GAGTGCACCA TATGCGGTGT

 16990 17000 17010 17020 17030 17040
 GAAATACCGC ACAGATGCGT AAGGAGAAAA TACCGCATCA GGCGCTCTTC CGCTTCCTCG

 17050 17060 17070 17080 17090 17100
 CTCACTGACT CGCTGCGCTC GGTCGTTCCG CTGGGGCGAG CGGTATCAGC TCACTCAAAG

 17110 17120 17130 17140 17150 17160
 GCGGTAATAC GGTTATCCAC AGAATCAGGG GATAACGCAG GAAAGAACAT GTGAGCAAA

 17170 17180 17190 17200 17210 17220
 GGCCAGCAAA AGGCCAGGAA CCGTAAAAAG GCCGGCGTTGC TGGCGTTTT CCATAGGCTC

 17230 17240 17250 17260 17270 17280
 CGCCCCCTG ACGAGCATCA CAAAATCGA CGCTCAAGTC AGAGGTGGCG AAACCCGACA

 17290 17300 17310 17320 17330 17340
 GGACTATAAA GATACCAGGC GTTTCCCCCT GGAAGCTCCC TCCTGCGCTC TCCTGTTCCG

 17350 17360 17370 17380 17390 17400
 ACCCTGCCGC TTACCGGATA CCTGTCCGCC TTTCTCCCTT CGGGAAAGCGT GGCGCTTTCT

 17410 17420 17430 17440 17450 17460
 CATAGCTCAC GCTGTAGGTA TCTCAGTTCG GTGTAGGTCG TTCGCTCCAA GCTGGGCTGT

 17470 17480 17490 17500 17510 17520
 GTGCACGAAC CCCCCGTTCA GCCCGACCGC TGCGCCTTAT CCGGTAACTA TCGTCTTGAG

 17530 17540 17550 17560 17570 17580
 TCCAACCCGG TAAGACACGA CTTATGCCA CTGGCAGCAG CCACTGGTAA CAGGATTAGC

 17590 17600 17610 17620 17630 17640
 AGAGCGAGGT ATGTAGGCGG TGCTACAGAG TTCTTGAAGT GGTGGCTAA CTACGGCTAC

 17650 17660 17670 17680 17690 17700
 ACTAGAAGGA CAGTATTTGG TATCTGCCT CTGCTGAAGC CAGTTACCTT CGGAAAAAGA

 17710 17720 17730 17740 17750 17760
 GTTGGTAGCT CTTGATCCGG CAAACAAACC ACCGCTGGTA GCGGTGGTTT TTTTGTGTTGC

 17770 17780 17790 17800 17810 17820
 AAGCAGCAGA TTACGCGCAG AAAAAAAGGA TCTCAAGAAG ATCCCTTGAT CTTTTCTACG

 17830 17840 17850 17860 17870 17880
 GGGTCTGACG CTCAGTGGAA CGAAAACCTCA CGTTAAGGGA TTTTGGTCAT GAGATTATCA

 17890 17900 17910 17920 17930 17940
 AAAAGGATCT TCACCTAGAT CCTTTAAAT TAAAAATGAA GTTTAAATC AATCTAAAGT

 17950 17960 17970 17980 17990 18000
 ATATATGAGT AAACCTGGTC TGACAGTTAC CAATGCTTAA TCAGTGAGGC ACCTATCTCA

 18010 18020 18030 18040 18050 18060
 GCGATCTGTC TATTCGTTTC ATCCATAGTT GCCTGACTCC CGTCGTGTA GATAACTACG

 18070 18080 18090 18100 18110 18120
 ATACGGGAGG GCTTACCATC TGGCCCCAGT GCTGCAATGA TACCGCGAGA CCCACGCTCA

 18130 18140 18150 18160 18170 18180
 CGGGCTCCAG ATTTATCAGC AATAAACAG CCAGCCGGAA GGGCCGAGCG CAGAAGTGGT

 18190 18200 18210 18220 18230 18240

DNASIS

Mandy + SE8N-SHL

51 / 51

CCTGCAACTT TATCCGCCTC CATCCAGTCT ATTAATTGTT GCCGGGAAGC TAGAGTAAGT
 18250 18260 18270 18280 18290 18300
 AGTTCGCCAG TTAATAGTTT GCGAACGTT GTGCCATTG CTGCAGGCAT CGTGGTGTCA
 18310 18320 18330 18340 18350 18360
 CGCTCGTCGT TTGGTATGGC TTCATTCAAGC TCCGGTTCCC AACGATCAAG GCGAGTTACA
 18370 18380 18390 18400 18410 18420
 TGATCCCCA TGTTGTCAA AAAAGCGGTT AGCTCCCTCG GTCCCTCGAT CGTTGTCAA
 18430 18440 18450 18460 18470 18480
 AGTAAGTTGG CCGCAGTGTG ATCACTCATG GTTATGGCAG CACTGCATAA TTCTCTTACT
 18490 18500 18510 18520 18530 18540
 GTCATGCCAT CCGTAAGATG CTTTCTGTG ACTGGTGAGT ACTCAACCAA GTCATTCTGA
 18550 18560 18570 18580 18590 18600
 GAATAGTGTG TGCGCGACC GAGTTGCTCT TGCCCGGCGT CAACACGGGA TAATACCGCG
 18610 18620 18630 18640 18650 18660
 CACATAGCA GAACCTTAAA AGTGCTCATC ATTGGAAAAC GTTCTCGGG GCGAAACTC
 18670 18680 18690 18700 18710 18720
 TCAAGGATCT TACCGCTGTT GAGATCCAGT TCGATGTAAC CCACTCGTGC ACCCAACTGA
 18730 18740 18750 18760 18770 18780
 TCTTCAGCAT CTTTACTTT CACCAGCGTT TCTGGGTGAG CAAAAACAGG AAGGCAAAAT
 18790 18800 18810 18820 18830 18840
 GCCGCAAAA AGGGATAAAG GGCGACACGG AAATGTTGAA TACTCATACT CTTCTTTTT
 18850 18860 18870 18880 18890 18900
 CAATATTATT GAAGCATTAA TCAGGGTTAT TGTCTCATGA GCGGATACAT ATTTGAATGT
 18910 18920 18930 18940 18950 18960
 ATTTAGAAA ATAAACAAAT AGGGGTCCG CGCACATTTC CCCGAAAAGT GCCACCTGAC
 18970 18980 18990 19000 19010 19020
 GTCTAAGAAA CCATTATTAT CATGACATTA ACCTATAAAA ATAGGCATAT CACGAGGCC
 19030 19040 19050 19060 19070 19080
 TTTCGTCTTC AAGAA.....

INTERNATIONAL SEARCH REPORT

Int. Application No
PCT/US 98/03935

A. CLASSIFICATION OF SUBJECT MATTER				
IPC 6	C12N15/90	C12N15/85	C12Q1/68	C12N5/
	C12N15/13	C07K16/28	C12N15/12	C07K14/705
	C12N15/62	C07K19/00		G01N33/53

According to International Patent Classification(IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C12Q C07K G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 94 11523 A (IDEC PHARMACEUTICALS CORPORATION (US); REFF MITCHELL E. (US)) 26 May 1994 cited in the application see abstract see page 9, line 21 - page 10, line 29 see page 41, line 19 - page 42, line 19; figure 6	1,4-8, 11,12, 25-29, 31,32
A	US 5 464 764 A (CAPECCHI MARIO R. AND KIRK THOMAS R.) 7 November 1995 see abstract see column 13, line 32 - column 14, line 5	1
A	WO 94 05784 A (UNITED STATES AMERICA REPRESENTED BY THE SECRETARY US DPT. AGRICULTURE) 17 March 1994 see abstract	1

	-/-	

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

23 July 1998

Date of mailing of the international search report

05/08/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.
Fax: (+31-70) 340-3016

Authorized officer

Macchia, G

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 98/03935

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 93 24642 A (TSI CORPORATION (US)) 9 December 1993 see abstract ----	1
A	BARNETT R.S. ET AL.: "Antibody production in chinese hamster ovary cells using an impaired selectable marker" ACS SYMPOSIUM SERIES: ANTIBODY EXPRESSION AND ENGINEERING, vol. 604, 1995, pages 27-40, XP002072464 -----	

INTERNATIONAL SEARCH REPORT

I. International Application No

PCT/US 98/03935

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9411523	A 26-05-1994	AU	682481 B	09-10-1997
		AU	5613294 A	08-06-1994
		CA	2149326 A	26-05-1994
		DE	669986 T	10-10-1996
		EP	0669986 A	06-09-1995
		ES	2088838 T	01-10-1996
		JP	8503138 T	09-04-1996
		US	5648267 A	15-07-1997
		US	5733779 A	31-03-1998
US 5464764	A 07-11-1995	US	5487992 A	30-01-1996
		US	5627059 A	06-05-1997
		US	5631153 A	20-05-1997
WO 9405784	A 17-03-1994	AU	4839493 A	29-03-1994
		MX	9305183 A	31-05-1994
WO 9324642	A 09-12-1993	AU	4401993 A	30-12-1993



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ :	A1	(11) International Publication Number: WO 98/41645 (43) International Publication Date: 24 September 1998 (24.09.98)
C12N 15/90, 15/85, C12Q 1/68, C12N 5/10, 9/12, 15/13, C07K 16/28, C12N 15/12, C07K 14/705, G01N 33/53, C12N 15/62, C07K 19/00		

(21) International Application Number: PCT/US98/03935	(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).
(22) International Filing Date: 9 March 1998 (09.03.98)	
(30) Priority Data:	
08/819,866 14 March 1997 (14.03.97) US	
09/023,715 13 February 1998 (13.02.98) US	
(71) Applicant: IDEC PHARMACEUTICALS CORPORATION [US/US]; 11011 Torreyana Road, San Diego, CA 92121 (US).	
(72) Inventors: REFF, Mitchell, E.; 4166 Combe Way, San Diego, CA 92122 (US). BARNETT, Richard, Spence; 306 Belmont Court, San Marcos, CA 92069 (US). McLACHLAN, Karen, Retta; Apartment B6, 766 South Nardo, Solana Beach, CA 92075 (US).	
(74) Agents: GESS, E., Joseph et al.; Burns, Doane, Swecker & Mathis L.L.P., P.O. Box 1404, Alexandria, VA 22313-1404 (US).	

(54) Title: METHOD FOR INTEGRATING GENES AT SPECIFIC SITES IN MAMMALIAN CELLS VIA HOMOLOGOUS RECOMBINATION AND VECTORS FOR ACCOMPLISHING THE SAME

(57) Abstract

A method for achieving site specific integration of a desired DNA at a target site in a mammalian cell via homologous recombination is described. This method provides for the reproducible selection of cell lines wherein a desired DNA is integrated at a predetermined transcriptionally active site previously marked with a marker plasmid. The method is particularly suitable for the production of mammalian cell lines which secrete mammalian proteins at high levels, in particular immunoglobulins. Vectors and vector combinations for use in the subject cloning method are also provided.

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		

Title of the Invention

METHOD FOR INTEGRATING GENES AT SPECIFIC SITES IN MAMMALIAN CELLS VIA HOMOLOGOUS RECOMBINATION AND VECTORS FOR ACCOMPLISHING THE SAME

5

Field of the Invention

The present invention relates to a process of targeting the integration of a desired exogenous DNA to a specific location within the genome of a mammalian cell. 10 More specifically, the invention describes a novel method for identifying a transcriptionally active target site ("hot spot") in the mammalian genome, and inserting a desired DNA at this site via homologous recombination. The invention also optionally provides the ability for 15 gene amplification of the desired DNA at this location by co-integrating an amplifiable selectable marker, e.g., DHFR, in combination with the exogenous DNA. The invention additionally describes the construction of novel vectors suitable for accomplishing the above, and 20 further provides mammalian cell lines produced by such methods which contain a desired exogenous DNA integrated at a target hot spot.

- 2 -

Background

Technology for expressing recombinant proteins in both prokaryotic and eukaryotic organisms is well established. Mammalian cells offer significant advantages over bacteria or yeast for protein production, resulting from their ability to correctly assemble, glycosylate and post-translationally modify recombinantly expressed proteins. After transfection into the host cells, recombinant expression constructs can be maintained as extrachromosomal elements, or may be integrated into the host cell genome. Generation of stably transfected mammalian cell lines usually involves the latter; a DNA construct encoding a gene of interest along with a drug resistance gene (dominant selectable marker) is introduced into the host cell, and subsequent growth in the presence of the drug allows for the selection of cells that have successfully integrated the exogenous DNA. In many instances, the gene of interest is linked to a drug resistant selectable marker which can later be subjected to gene amplification. The gene encoding dihydrofolate reductase (DHFR) is most commonly used for this purpose. Growth of cells in the presence of methotrexate, a competitive inhibitor of DHFR, leads to increased DHFR production by means of amplification of the DHFR gene. As flanking regions of DNA will also become amplified, the resultant coamplification of a DHFR linked gene in the transfected cell line can lead to increased protein

- 3 -

production, thereby resulting in high level expression of the gene of interest.

While this approach has proven successful, there are a number of problems with the system because of the random nature of the integration event. These problems exist because expression levels are greatly influenced by the effects of the local genetic environment at the gene locus, a phenomena well documented in the literature and generally referred to as "position effects" (for example, see Al-Shawi et al, *Mol. Cell. Biol.*, 10:1192-1198 (1990); Yoshimura et al, *Mol. Cell. Biol.*, 7:1296-1299 (1987)). As the vast majority of mammalian DNA is in a transcriptionally inactive state, random integration methods offer no control over the transcriptional fate of the integrated DNA. Consequently, wide variations in the expression level of integrated genes can occur, depending on the site of integration. For example, integration of exogenous DNA into inactive, or transcriptionally "silent" regions of the genome will result in little or no expression. By contrast integration into a transcriptionally active site may result in high expression.

Therefore, when the goal of the work is to obtain a high level of gene expression, as is typically the desired outcome of genetic engineering methods, it is generally necessary to screen large numbers of transfec-tants to find such a high producing clone.

- 4 -

Additionally, random integration of exogenous DNA into the genome can in some instances disrupt important cellular genes, resulting in an altered phenotype.

These factors can make the generation of high expressing stable mammalian cell lines a complicated and laborious process.

Recently, our laboratory has described the use of DNA vectors containing translationally impaired dominant selectable markers in mammalian gene expression. (This is disclosed in U.S. Serial No. 08/147,696 filed November 3, 1993, recently allowed).

These vectors contain a translationally impaired neomycin phosphotransferase (neo) gene as the dominant selectable marker, artificially engineered to contain an intron into which a DHFR gene along with a gene or genes of interest is inserted. Use of these vectors as expression constructs has been found to significantly reduce the total number of drug resistant colonies produced, thereby facilitating the screening procedure in relation to conventional mammalian expression vectors.

Furthermore, a significant percentage of the clones obtained using this system are high expressing clones.

These results are apparently attributable to the modifications made to the neo selectable marker. Due to the translational impairment of the neo gene, transfected cells will not produce enough neo protein to survive drug selection, thereby decreasing the overall

- 5 -

number of drug resistant colonies. Additionally, a higher percentage of the surviving clones will contain the expression vector integrated into sites in the genome where basal transcription levels are high,
5 resulting in overproduction of neo, thereby allowing the cells to overcome the impairment of the neo gene. Concomitantly, the genes of interest linked to neo will be subject to similar elevated levels of transcription. This same advantage is also true as a result of the
10 artificial intron created within neo; survival is dependent on the synthesis of a functional neo gene, which is in turn dependent on correct and efficient splicing of the neo introns. Moreover, these criteria are more likely to be met if the vector DNA has
15 integrated into a region which is already highly transcriptionally active.

Following integration of the vector into a transcriptionally active region, gene amplification is performed by selection for the DHFR gene. Using this system,
20 it has been possible to obtain clones selected using low levels of methotrexate (50nM), containing few (<10) copies of the vector which secrete high levels of protein (>55pg/cell/day). Furthermore, this can be achieved in a relatively short period of time. However,
25 the success in amplification is variable. Some transcriptionally active sites cannot be amplified and

- 6 -

therefore the frequency and extent of amplification from a particular site is not predictable.

Overall, the use of these translationally impaired vectors represents a significant improvement over other methods of random integration. However, as discussed, the problem of lack of control over the integration site remains a significant concern.

One approach to overcome the problems of random integration is by means of gene targeting, whereby the exogenous DNA is directed to a specific locus within the host genome. The exogenous DNA is inserted by means of homologous recombination occurring between sequences of DNA in the expression vector and the corresponding homologous sequence in the genome. However, while this type of recombination occurs at a high frequency naturally in yeast and other fungal organisms, in higher eukaryotic organisms it is an extremely rare event. In mammalian cells, the frequency of homologous versus non-homologous (random integration) recombination is reported to range from 1/100 to 1/5000 (for example, see Capecchi, *Science*, 244:1288-1292 (1989); Morrow and Kucherlapati, *Curr. Op. Biotech.*, 4:577-582 (1993)).

One of the earliest reports describing homologous recombination in mammalian cells comprised an artificial system created in mouse fibroblasts (Thomas et al, *Cell*, 44:419-428 (1986)). A cell line containing a mutated, non-functional version of the neo gene integrated into

- 7 -

the host genome was created, and subsequently targeted with a second non-functional copy of neo containing a different mutation. Reconstruction of a functional neo gene could occur only by gene targeting. Homologous 5 recombinants were identified by selecting for G418 resistant cells, and confirmed by analysis of genomic DNA isolated from the resistant clones.

Recently, the use of homologous recombination to replace the heavy and light immunoglobulin genes at 10 endogenous loci in antibody secreting cells has been reported. (U.S. Patent No. 5,202,238, Fell et al, (1993).) However, this particular approach is not widely applicable, because it is limited to the production of immunoglobulins in cells which 15 endogenously express immunoglobulins, e.g., B cells and myeloma cells. Also, expression is limited to single copy gene levels because co-amplification after homologous recombination is not included. The method is further complicated by the fact that two separate 20 integration events are required to produce a functional immunoglobulin: one for the light chain gene followed by one for the heavy chain gene.

An additional example of this type of system has 25 been reported in NS/0 cells, where recombinant immunoglobulins are expressed by homologous recombination into the immunoglobulin gamma 2A locus (Hollis et al, international patent application #

PCT/IB95 (00014).) Expression levels obtained from this site were extremely high - on the order of 20pg/cell/day from a single copy integrant. However, as in the above example, expression is limited to this level because an amplifiable gene is not cointegrated in this system.

5 Also, other researchers have reported aberrant glycosylation of recombinant proteins expressed in NS/0 cells (for example, see Flesher et al, *Biotech. and Bioeng.*, 48:399-407 (1995)), thereby limiting the applicability of this approach.

10 The cre-loxP recombination system from bacteriophage P1 has recently been adapted and used as a means of gene targeting in eukaryotic cells.

15 Specifically, the site specific integration of exogenous DNA into the Chinese hamster ovary (CHO) cell genome using cre recombinase and a series of lox containing vectors have been described. (Fukushige and Sauer, *Proc. Natl. Acad. Sci. USA*, 89:7905-7909 (1992).) This system is attractive in that it provides for

20 reproducible expression at the same chromosomal location. However, no effort was made to identify a chromosomal site from which gene expression is optimal, and as in the above example, expression is limited to single copy levels in this system. Also, it is

25 complicated by the fact that one needs to provide for expression of a functional recombinase enzyme in the mammalian cell.

The use of homologous recombination between an introduced DNA sequence and its endogenous chromosomal locus has also been reported to provide a useful means of genetic manipulation in mammalian cells, as well as 5 in yeast cells. (See e.g., Bradley et al, *Meth. Enzymol.*, 223:855-879 (1993); Capecchi, *Science*, 244:1288-1292 (1989); Rothstein et al, *Meth. Enzymol.*, 194:281-301 (1991)). To date, most mammalian gene targeting studies have been directed toward gene 10 disruption ("knockout") or site-specific mutagenesis of selected target gene loci in mouse embryonic stem (ES) cells. The creation of these "knockout" mouse models has enabled scientists to examine specific structure-function issues and examine the biological 15 importance of a myriad of mouse genes. This field of research also has important implications in terms of potential gene therapy applications.

Also, vectors have recently been reported by Cell-tech (Kent, U.K.) which purportedly are targeted to 20 transcriptionally active sites in NSO cells, which do not require gene amplification (Peakman et al, *Hum. Antibod. Hybridomas*, 5:65-74 (1994)). However, levels of immunoglobulin secretion in these unamplified cells have not been reported to exceed 20pg/cell/day, while in 25 amplified CHO cells, levels as high as 100pg/cell/day can be obtained (Id.).

- 10 -

It would be highly desirable to develop a gene targeting system which reproducibly provided for the integration of exogenous DNA into a predetermined site in the genome known to be transcriptionally active.

5 Also, it would be desirable if such a gene targeting system would further facilitate co-amplification of the inserted DNA after integration. The design of such a system would allow for the reproducible and high level expression of any cloned gene of interest in a mammalian 10 cell, and undoubtedly would be of significant interest to many researchers.

In this application, we provide a novel mammalian expression system, based on homologous recombination occurring between two artificial substrates contained in 15 two different vectors. Specifically, this system uses a combination of two novel mammalian expression vectors, referred to as a "marking" vector and a "targeting" vector.

Essentially, the marking vector enables the identification and marking of a site in the mammalian genome 20 which is transcriptionally active, i.e., a site at which gene expression levels are high. This site can be regarded as a "hot spot" in the genome. After integration of the marking vector, the subject expression system 25 enables another DNA to be integrated at this site, i.e., the targeting vector, by means of homologous recombination occurring between DNA sequences common to

- 11 -

both vectors. This system affords significant advantages over other homologous recombination systems.

Unlike most other homologous systems employed in mammalian cells, this system exhibits no background.

5 Therefore, cells which have only undergone random integration of the vector do not survive the selection. Thus, any gene of interest cloned into the targeting plasmid is expressed at high levels from the marked hot spot. Accordingly, the subject method of gene expression substantially or completely eliminates the problems inherent to systems of random integration, discussed in detail above. Moreover, this system provides reproducible and high level expression of any recombinant protein at the same transcriptionally active site in the 10 mammalian genome. In addition, gene amplification may be effected at this particular transcriptionally active site by including an amplifiable dominant selectable marker (e.g. DHFR) as part of the marking vector.

Objects of the Invention

20 Thus, it is an object of the invention to provide an improved method for targeting a desired DNA to a specific site in a mammalian cell.

It is a more specific object of the invention to provide a novel method for targeting a desired DNA to a 25 specific site in a mammalian cell via homologous recombination.

- 12 -

It is another specific object of the invention to provide novel vectors for achieving site specific integration of a desired DNA in a mammalian cell.

5 It is still another object of the invention to provide novel mammalian cell lines which contain a desired DNA integrated at a predetermined site which provides for high expression.

10 It is a more specific object of the invention to provide a novel method for achieving site specific integration of a desired DNA in a Chinese hamster ovary (CHO) cell.

15 It is another more specific object of the invention to provide a novel method for integrating immunoglobulin genes, or any other genes, in mammalian cells at predetermined chromosomal sites that provide for high expression.

20 It is another specific object of the invention to provide novel vectors and vector combinations suitable for integrating immunoglobulin genes into mammalian cells at predetermined sites that provide for high expression.

25 It is another object of the invention to provide mammalian cell lines which contain immunoglobulin genes integrated at predetermined sites that provide for high expression.

It is an even more specific object of the invention to provide a novel method for integrating immunoglobulin

genes into CHO cells that provide for high expression, as well as novel vectors and vector combinations that provide for such integration of immunoglobulin genes into CHO cells.

5 In addition, it is a specific object of the invention to provide novel CHO cell lines which contain immunoglobulin genes integrated at predetermined sites that provide for high expression, and have been amplified by methotrexate selection to secrete even greater amounts
10 of functional immunoglobulins.

Brief Description of the Figures

Figure 1 depicts a map of a marking plasmid according to the invention referred to as Desmond. The plasmid is shown in circular form (1a) as well as a
15 linearized version used for transfection (1b).

Figure 2(a) shows a map of a targeting plasmid referred to "Molly". Molly is shown here encoding the anti-CD20 immunoglobulin genes, expression of which is described in Example 1.

20 Figure 2(b) shows a linearized version of Molly, after digestion with the restriction enzymes *Kpn*1 and *Pac*1. This linearized form was used for transfection.

Figure 3 depicts the potential alignment between Desmond sequences integrated into the CHO genome, and
25 incoming targeting Molly sequences. One potential ar-

- 14 -

rangement of Molly integrated into Desmond after homologous recombination is also presented.

Figure 4 shows a Southern analysis of single copy Desmond clones. Samples are as follows:

- 5 Lane 1: λ HindIII DNA size marker
- Lane 2: Desmond clone 10F3
- Lane 3: Desmond clone 10C12
- Lane 4: Desmond clone 15C9
- Lane 5: Desmond clone 14B5
- 10 Lane 6: Desmond clone 9B2

Figure 5 shows a Northern analysis of single copy Desmond clones. Samples are as follows: Panel A: northern probed with CAD and DHFR probes, as indicated on the figure. Panel B: duplicate northern, probed with CAD and HisD probes, as indicated. The RNA samples loaded in panels A and B are as follows:

- Lane 1: clone 9B2, lane 2; clone 10C12, lane 3; clone 14B5, lane 4; clone 15C9, lane 5; control RNA from CHO transfected with a HisD and DHFR containing plasmid,
- 20 lane 6; untransfected CHO.

Figure 6 shows a Southern analysis of clones resulting from the homologous integration of Molly into Desmond. Samples are as follows:

- Lane 1: λ HindIII DNA size markers, Lane 2: 20F4, lane 3; 25
- 25 5F9, lane 4; 21C7, lane 5; 24G2, lane 6; 25E1, lane 7; 28C9, lane 8; 29F9, lane 9; 39G11, lane 10; 42F9, lane 11; 50G10, lane 12; Molly plasmid DNA, linearized with

- 15 -

BglII (top band) and cut with BglII and KpnI (lower band), lane 13; untransfected Desmond.

Figures 7A through 7G contain the Sequence Listing for Desmond.

5 Figures 8A through 8I contain the Sequence Listing for Molly-containing anti-CD20.

Figure 9 contains a map of the targeting plasmid, "Mandy," shown here encoding anti-CD23 genes, the expression of which is disclosed in Example 5.

10 Figures 10A through 10N contain the sequence listing of "Mandy" containing the anti-CD23 genes as disclosed in Example 5.

Detailed Description of the Invention

The invention provides a novel method for integrating a desired exogenous DNA at a target site within the genome of a mammalian cell via homologous recombination. Also, the invention provides novel vectors for achieving the site specific integration of a DNA at a target site in the genome of a mammalian cell.

20 More specifically, the subject cloning method provides for site specific integration of a desired DNA in a mammalian cell by transfection of such cell with a "marker plasmid" which contains a unique sequence that is foreign to the mammalian cell genome and which 25 provides a substrate for homologous recombination, followed by transfection with a "target plasmid" containing

- 16 -

a sequence which provides for homologous recombination with the unique sequence contained in the marker plasmid, and further comprising a desired DNA that is to be integrated into the mammalian cell. Typically, the 5 integrated DNA will encode a protein of interest, such as an immunoglobulin or other secreted mammalian glycoprotein.

The exemplified homologous recombination system uses the neomycin phosphotransferase gene as a dominant 10 selectable marker. This particular marker was utilized based on the following previously published observations;

(i) the demonstrated ability to target and restore function to a mutated version of the neo gene (cited 15 earlier) and

(ii) our development of translationally impaired expression vectors, in which the neo gene has been artificially created as two exons with a gene of interest inserted in the intervening intron; neo exons are correctly spliced and translated in vivo, producing a functional protein and thereby conferring G418 resistance on the resultant cell population. In this application, the neo gene is split into three exons. The third exon of 20 neo is present on the "marker" plasmid and becomes integrated into the host cell genome upon integration of the 25 marker plasmid into the mammalian cells. Exons 1 and 2 are present on the targeting plasmid, and are separated

by an intervening intron into which at least one gene of interest is cloned. Homologous recombination of the targeting vector with the integrated marking vector results in correct splicing of all three exons of the 5 neo gene and thereby expression of a functional neo protein (as determined by selection for G418 resistant colonies). Prior to designing the current expression system, we had experimentally tested the functionality of such a triply spliced neo construct in mammalian 10 cells. The results of this control experiment indicated that all three neo exons were properly spliced and therefore suggested the feasibility of the subject invention.

However, while the present invention is exemplified 15 using the neo gene, and more specifically a triple split neo gene, the general methodology should be efficacious with other dominant selectable markers.

As discussed in greater detail *infra*, the present invention affords numerous advantages to conventional 20 gene expression methods, including both random integration and gene targeting methods. Specifically, the subject invention provides a method which reproducibly allows for site-specific integration of a desired DNA into a transcriptionally active domain of a mammalian 25 cell. Moreover, because the subject method introduces an artificial region of "homology" which acts as a unique substrate for homologous recombination and the

insertion of a desired DNA, the efficacy of subject invention does not require that the cell endogenously contain or express a specific DNA. Thus, the method is generically applicable to all mammalian cells, and can
5 be used to express any type of recombinant protein.

The use of a triply spliced selectable marker, e.g., the exemplified triply spliced neo construct, guarantees that all G418 resistant colonies produced will arise from a homologous recombination event (random integrants will not produce a functional neo gene and consequently will not survive G418 selection). Thus,
10 the subject invention makes it easy to screen for the desired homologous event. Furthermore, the frequency of additional random integrations in a cell that has undergone a homologous recombination event appears to be low.
15

Based on the foregoing, it is apparent that a significant advantage of the invention is that it substantially reduces the number of colonies that need be screened to identify high producer clones, i.e., cell
20 lines containing a desired DNA which secrete the corresponding protein at high levels. On average, clones containing integrated desired DNA may be identified by screening about 5 to 20 colonies (compared to several thousand which must be screened when using standard
25 random integration techniques, or several hundred using the previously described intronic insertion vectors) Additionally, as the site of integration was preselected

- 19 -

and comprises a transcriptionally active domain, all exogenous DNA expressed at this site should produce comparable, i.e. high levels of the protein of interest.

Moreover, the subject invention is further advantageous in that it enables an amplifiable gene to be inserted on integration of the marking vector. Thus, when a desired gene is targeted to this site via homologous recombination, the subject invention allows for expression of the gene to be further enhanced by gene amplification. In this regard, it has been reported in from the literature that different genomic sites have different capacities for gene amplification (Meinkoth et al, *Mol. Cell Biol.*, 7:1415-1424 (1987)). Therefore, this technique is further advantageous as it allows for the placement of a desired gene of interest at a specific site that is both transcriptionally active and easily amplified. Therefore, this should significantly reduce the amount of time required to isolate such high producers.

Specifically, while conventional methods for the construction of high expressing mammalian cell lines can take 6 to 9 months, the present invention allows for such clones to be isolated on average after only about 3-6 months. This is due to the fact that conventionally isolated clones typically must be subjected to at least three rounds of drug resistant gene amplification in order to reach satisfactory levels of gene expression.

- 20 -

As the homologously produced clones are generated from a preselected site which is a high expression site, fewer rounds of amplification should be required before reaching a satisfactory level of production.

5 Still further, the subject invention enables the reproducible selection of high producer clones wherein the vector is integrated at low copy number, typically single copy. This is advantageous as it enhances the stability of the clones and avoids other potential adverse side-effects associated with high copy number. As described *supra*, the subject homologous recombination system uses the combination of a "marker plasmid" and a "targeting plasmid" which are described in more detail below.

15 The "marker plasmid" which is used to mark and identify a transcriptionally hot spot will comprise at least the following sequences:

(i) a region of DNA that is heterologous or unique to the genome of the mammalian cell, which functions as a source of homology, allows for homologous recombination (with a DNA contained in a second target plasmid). More specifically, the unique region of DNA (i) will generally comprise a bacterial, viral, yeast synthetic, or other DNA which is not normally present in the 25 mammalian cell genome and which further does not comprise significant homology or sequence identity to DNA contained in the genome of the mammalian cell.

Essentially, this sequence should be sufficiently different to mammalian DNA that it will not significantly recombine with the host cell genome via homologous recombination. The size of such unique DNA 5 will generally be at least about 2 to 10 kilobases in size, or higher, more preferably at least about 10kb, as several other investigators have noted an increased frequency of targeted recombination as the size of the homology region is increased (Capecchi, *Science*, 10 244:1288-1292 (1989)).

The upper size limit of the unique DNA which acts as a site for homologous recombination with a sequence in the second target vector is largely dictated by potential stability constraints (if DNA is too large it 15 may not be easily integrated into a chromosome and the difficulties in working with very large DNAs.

(ii) a DNA including a fragment of a selectable marker DNA, typically an exon of a dominant selectable marker gene. The only essential feature of this DNA is 20 that it not encode a functional selectable marker protein unless it is expressed in association with a sequence contained in the target plasmid. Typically, the target plasmid will comprise the remaining exons of the dominant selectable marker gene (those not comprised in 25 "targeting" plasmid). Essentially, a functional selectable marker should only be produced if homologous recombination occurs (resulting in the association and

expression of this marker DNA (i) sequence together with the portion(s) of the selectable marker DNA fragment which is (are) contained in the target plasmid).

As noted, the current invention exemplifies the
5 use of the neomycin phosphotransferase gene as the dominant selectable marker which is "split" in the two vectors. However, other selectable markers should also be suitable, e.g., the *Salmonella histidinol dehydrogenase* gene, *hygromycin phosphotransferase* gene, *herpes simplex* virus thymidine kinase gene, adenosine deaminase gene,
10 glutamine synthetase gene and hypoxanthine-guanine phosphoribosyl transferase gene.

(iii) a DNA which encodes a functional selectable marker protein, which selectable marker is different
15 from the selectable marker DNA (ii). This selectable marker provides for the successful selection of mammalian cells wherein the marker plasmid is successfully integrated into the cellular DNA. More preferably, it is desirable that the marker plasmid comprise two such
20 dominant selectable marker DNAs, situated at opposite ends of the vector. This is advantageous as it enables integrants to be selected using different selection agents and further enables cells which contain the entire vector to be selected. Additionally, one marker
25 can be an amplifiable marker to facilitate gene amplification as discussed previously. Any of the

dominant selectable marker listed in (ii) can be used as well as others generally known in the art.

Moreover, the marker plasmid may optionally further comprise a rare endonuclease restriction site. This is potentially desirable as this may facilitate cleavage. If present, such rare restriction site should be situated close to the middle of the unique region that acts as a substrate for homologous recombination. Preferably such sequence will be at least about 12 nucleotides.

The introduction of a double stranded break by similar methodology has been reported to enhance the frequency of homologous recombination. (Choulika et al, *Mol. Cell. Biol.*, 15:1968-1973 (1995)). However, the presence of such sequence is not essential.

The "targeting plasmid" will comprise at least the following sequences:

(1) the same unique region of DNA contained in the marker plasmid or one having sufficient homology or sequence identity therewith that said DNA is capable of combining via homologous recombination with the unique region (i) in the marker plasmid. Suitable types of DNAs are described *supra* in the description of the unique region of DNA (1) in the marker plasmid.

(2) The remaining exons of the dominant selectable marker, one exon of which is included as (ii) in the marker plasmid listed above. The essential features of this DNA fragment is that it result in a functional

(selectable) marker protein only if the target plasmid integrates via homologous recombination (wherein such recombination results in the association of this DNA with the other fragment of the selectable marker DNA contained in the marker plasmid) and further that it allow for insertion of a desired exogenous DNA. Typically, this DNA will comprise the remaining exons of the selectable marker DNA which are separated by an intron. For example, this DNA may comprise the first two exons of the neo gene and the marker plasmid may comprise the third exon (back third of neo).

(3) The target plasmid will also comprise a desired DNA, e.g., one encoding a desired polypeptide, preferably inserted within the selectable marker DNA fragment contained in the plasmid. Typically, the DNA will be inserted in an intron which is comprised between the exons of the selectable marker DNA. This ensures that the desired DNA is also integrated if homologous recombination of the target plasmid and the marker plasmid occurs. This intron may be naturally occurring or it may be engineered into the dominant selectable marker DNA fragment.

This DNA will encode any desired protein, preferably one having pharmaceutical or other desirable properties. Most typically the DNA will encode a mammalian protein, and in the current examples provided, an immunoglobulin or an immunoadhesin. However the

invention is not in any way limited to the production of immunoglobulins.

As discussed previously, the subject cloning method is suitable for any mammalian cell as it does not require for efficacy that any specific mammalian sequence or sequences be present. In general, such mammalian cells will comprise those typically used for protein expression, e.g., CHO cells, myeloma cells, COS cells, BHK cells, Sp2/0 cells, NIH 3T3 and HeLa cells. In the examples which follow, CHO cells were utilized. The advantages thereof include the availability of suitable growth medium, their ability to grow efficiently and to high density in culture, and their ability to express mammalian proteins such as immunoglobulins in biologically active form.

Further, CHO cells were selected in large part because of previous usage of such cells by the inventors for the expression of immunoglobulins (using the translationally impaired dominant selectable marker containing vectors described previously). Thus, the present laboratory has considerable experience in using such cells for expression. However, based on the examples which follow, it is reasonable to expect similar results will be obtained with other mammalian cells.

In general, transformation or transfection of mammalian cells according to the subject invention will be effected according to conventional methods. So that the

invention may be better understood, the construction of exemplary vectors and their usage in producing integrants is described in the examples below.

EXAMPLE 1

5

Design and Preparation of Marker
and Targeting Plasmid DNA Vectors

The marker plasmid herein referred to as "Desmond" was assembled from the following DNA elements:

(a) Murine dihydrofolate reductase gene (DHFR),
10 incorporated into a transcription cassette, comprising the mouse beta globin promoter 5" to the DHFR start site, and bovine growth hormone poly adenylation signal 3" to the stop codon. The DHFR transcriptional cassette was isolated from TCAE6, an expression vector created
15 previously in this laboratory (Newman et al, 1992, *Bio-technology*, 10:1455-1460).

(b) E. coli β-galactosidase gene - commercially available, obtained from Promega as pSV-β-galactosidase control vector, catalog # E1081.

20 (c) Baculovirus DNA, commercially available, purchased from Clontech as pBAKPAK8, cat # 6145-1.

(d) Cassette comprising promoter and enhancer elements from Cytomegalovirus and SV40 virus. The cassette was generated by PCR using a derivative of expression
25 vector TCAE8 (Reff et al, *Blood*, 83:435-445 (1994)). The enhancer cassette was inserted within the baculo-

- 27 -

virus sequence, which was first modified by the insertion of a multiple cloning site.

(e) E. coli GUS (glucuronidase) gene, commercially available, purchased from Clontech as pB101, cat. #

5 6017-1.

(f) Firefly luciferase gene, commercially available, obtained from Promega as pGEM-Luc (catalog # E1541).

(g) S. typhimurium histidinol dehydrogenase gene (HisD). This gene was originally a gift from (Donahue et al, *Gene*, 18:47-59 (1982)), and has subsequently been incorporated into a transcription cassette comprising the mouse beta globin major promoter 5' to the gene, and the SV40 polyadenylation signal 3' to the gene.

15 The DNA elements described in (a)-(g) were combined into a pBR derived plasmid backbone to produce a 7.7kb contiguous stretch of DNA referred to in the attached figures as "homology". Homology in this sense refers to sequences of DNA which are not part of the mammalian genome and are used to promote homologous recombination between transfected plasmids sharing the same homology DNA sequences.

20 (h) Neomycin phosphotransferase gene from TN5 (Davis and Smith, *Ann. Rev. Micro.*, 32:469-518 (1978)).

25 The complete neo gene was subcloned into pBluescript SK- (Stratagene catalog # 212205) to facilitate genetic manipulation. A synthetic linker was then inserted into

a unique PstI site occurring across the codons for amino acid 51 and 52 of neo. This linker encoded the necessary DNA elements to create an artificial splice donor site, intervening intron and splice acceptor site within the neo gene, thus creating two separate exons, presently referred to as neo exon 1 and 2. Neo exon 1 encodes the first 51 amino acids of neo, while exon 2 encodes the remaining 203 amino acids plus the stop codon of the protein A NotI cloning site was also created within the 5 intron.

Neo exon 2 was further subdivided to produce neo exons 2 and 3. This was achieved as follows: A set of PCR primers were designed to amplify a region of DNA encoding neo exon 1, intron and the first 111 2/3 amino acids of exon2. The 3' PCR primer resulted in the introduction of a new 5' splice site immediately after the second nucleotide of the codon for amino acid 111 in exon 2, therefore generating a new smaller exon 2. The DNA fragment now encoding the original exon 1, intron and new exon 2 was then subcloned and propagated in a pBR based vector. The remainder of the original exon 2 was used as a template for another round of PCR amplification, which generated "exon3". The 5' primer for this round of amplification introduced a new splice acceptor site at the 5' side of the newly created exon 25 3, i.e. before the final nucleotide of the codon for amino acid 111. The resultant 3 exons of neo encode the

following information: exon 1 - the first 51 amino acids of neo; exon 2 - the next 111 2/3 amino acids, and exon 3 the final 91 1/3 amino acids plus the translational stop codon of the neo gene.

5 Neo exon 3 was incorporated along with the above mentioned DNA elements into the marking plasmid "Desmond". Neo exons 1 and 2 were incorporated into the targeting plasmid "Molly". The NotI cloning site created within the intron between exons 1 and 2 was used in 10 subsequent cloning steps to insert genes of interest into the targeting plasmid.

A second targeting plasmid "Mandy" was also generated. This plasmid is almost identical to "Molly" (some restriction sites on the vector have been changed) 15 except that the original HisD and DHFR genes contained in "Molly" were inactivated. These changes were incorporated because the Desmond cell line was no longer being cultured in the presence of Histidinol, therefore it seemed unnecessary to include a second copy of the 20 HisD gene. Additionally, the DHFR gene was inactivated to ensure that only a single DHFR gene, namely the one present in the Desmond marked site, would be amplifiable in any resulting cell lines. "Mandy" was derived from "Molly" by the following modifications:

25 (i) A synthetic linker was inserted in the middle of the DHFR coding region. This linker created a stop codon and shifted the remainder of the DHFR coding

- 30 -

region out of frame, therefore rendering the gene nonfunctional.

(ii) A portion of the HisD gene was deleted and replaced with a PCR generated HisD fragment lacking the 5 promoter and start codon of the gene.

Figure 1 depicts the arrangement of these DNA elements in the marker plasmid "Desmond". Figure 2 depicts the arrangement of these elements in the first targeting plasmid, "Molly". Figure 3 illustrates the possible 10 arrangement in the CHO genome, of the various DNA elements after targeting and integration of Molly DNA into Desmond marked CHO cells. Figure 9 depicts the targeting plasmid "Mandy."

Construction of the marking and targeting plasmids 15 from the above listed DNA elements was carried out following conventional cloning techniques (see, e.g., Molecular Cloning, A Laboratory Manual, J. Sambrook et al, 1987, Cold Spring Harbor Laboratory Press, and Current Protocols in Molecular Biology, F. M. Ausubel et 20 al, eds., 1987, John Wiley and Sons). All plasmids were propagated and maintained in E. coli XLI blue (Stratagene, cat. # 200236). Large scale plasmid preparations were prepared using Promega Wizard Maxiprep DNA Purification System®, according to the 25 manufacturer's directions.

- 31 -

EXAMPLE 2

Construction of a Marked CHO Cell Line

1. Cell Culture and Transfection Procedures to
Produced Marked CHO Cell Line

5 Marker plasmid DNA was linearized by digestion
overnight at 37°C with Bst1107I. Linearized vector was
ethanol precipitated and resuspended in sterile TE to a
concentration of 1mg/ml. Linearized vector was intro-
duced into DHFR-Chinese hamster ovary cells (CHO cells)
10 DG44 cells (Urlaub et al, *Som. Cell and Mol. Gen.*,
12:555-566 (1986)) by electroporation as follows.

Exponentially growing cells were harvested by cen-
trifugation, washed once in ice cold SBS (sucrose
buffered solution, 272mM sucrose, 7mM sodium phosphate,
15 pH 7.4, 1mM magnesium chloride) then resuspended in SBS
to a concentration of 10⁷ cells/ml. After a 15 minute
incubation on ice, 0.4ml of the cell suspension was
mixed with 40μg linearized DNA in a disposable
electroporation cuvette. Cells were shocked using a BTX
20 electrocell manipulator (San Diego, CA) set at 230
volts, 400 microfaraday capacitance, 13 ohm resistance.
Shocked cells were then mixed with 20 ml of prewarmed
CHO growth media (CHO-S-SFMII, Gibco/BRL, catalog #
31033-012) and plated in 96 well tissue culture plates.
25 Forty eight hours after electroporation, plates were fed
with selection media (in the case of transfection with
Desmond, selection media is CHO-S-SFMII without

- 32 -

hypoxanthine or thymidine, supplemented with 2mM Histidinol (Sigma catalog # H6647)). Plates were maintained in selection media for up to 30 days, or until some of the wells exhibited cell growth. These cells
5 were then removed from the 96 well plates and expanded ultimately to 120 ml spinner flasks where they were maintained in selection media at all times.

EXAMPLE 3

Characterization of Marked CHO Cell Lines

10 (a) Southern Analysis

Genomic DNA was isolated from all stably growing Desmond marked CHO cells. DNA was isolated using the Invitrogen Easy[®] DNA kit, according to the manufacturer's directions. Genomic DNA was then digested with
15 HindIII overnight at 37°C, and subjected to Southern analysis using a PCR generated digoxigenin labelled probe specific to the DHFR gene. Hybridizations and washes were carried out using Boehringer Mannheim's DIG easy hyb (catalog # 1603 558) and DIG Wash and Block
20 Buffer Set (catalog # 1585 762) according to the manufacturer's directions. DNA samples containing a single band hybridizing to the DHFR probe were assumed to be Desmond clones arising from a single cell which had integrated a single copy of the plasmid. These clones
25 were retained for further analysis. Out of a total of 45 HisD resistant cell lines isolated, only 5 were

- 33 -

single copy integrants. Figure 4 shows a Southern blot containing all 5 of these single copy Desmond clones. Clone names are provided in the figure legend.

(b) Northern Analysis

5 Total RNA was isolated from all single copy Desmond clones using TRIzol reagent (Gibco/BRL cat # 15596-026) according to the manufacturer's directions. 10-20 μ g RNA from each clone was analyzed on duplicate formaldehyde gels. The resulting blots were probed with PCR generated digoxigenin labelled DNA probes to (i) DHFR message, (ii) HisD message and (iii) CAD message. CAD is a trifunctional protein involved in uridine biosynthesis (Wahl et al, *J. Biol. Chem.*, 254, 17:8679-8689 (1979)), and is expressed equally in all cell types. It is used here as an internal control to help quantitate RNA loading. Hybridizations and washes were carried out using the above mentioned Boehringer Mannheim reagents. The results of the Northern analysis are shown in Figure 5. The single copy Desmond clone exhibiting the highest levels of both the His D and DHFR message is clone 15C9, shown in lane 4 in both panels of the figure. This clone was designated as the "marked cell line" and used in future targeting experiments in CHO, examples of which are presented in the following sections.

- 34 -

EXAMPLE 4

Expression of Anti-CD20 Antibody
in Desmond Marked CHO Cells

C2B8, a chimeric antibody which recognizes B-cell

5 surface antigen CD20, has been cloned and expressed previously in our laboratory. (Reff et al, *Blood*, 83:434-45 (1994)). A 4.1 kb DNA fragment comprising the C2B8 light and heavy chain genes, along with the necessary regulatory elements (eukaryotic promoter and polyadenylation signals) was inserted into the artificial 10 intron created between exons 1 and 2 of the neo gene contained in a pBR derived cloning vector. This newly generated 5kb DNA fragment (comprising neo exon 1, C2B8 and neo exon 2) was excised and used to assemble the targeting plasmid Molly. The other DNA elements used in 15 the construction of Molly are identical to those used to construct the marking plasmid Desmond, identified previously. A complete map of Molly is shown in Fig. 2.

The targeting vector Molly was linearized prior to 20 transfection by digestion with *Kpn*1 and *Pac*1, ethanol precipitated and resuspended in sterile TE to a concentration of 1.5mg/mL. Linearized plasmid was introduced into exponentially growing Desmond marked cells essentially as described, except that 80 μ g DNA was used in 25 each electroporation. Forty eight hours postelectroporation, 96 well plates were supplemented with selection medium - CHO-SSFMII supplemented with 400 μ g/mL Geneti-

cin (G418, Gibco/BRL catalog # 10131-019). Plates were maintained in selection medium for up to 30 days, or until cell growth occurred in some of the wells. Such growth was assumed to be the result of clonal expansion 5 of a single G418 resistant cell. The supernatants from all G418 resistant wells were assayed for C2B8 production by standard ELISA techniques, and all productive clones were eventually expanded to 120mL spinner flasks and further analyzed.

10 Characterization of Antibody secreting Targeted Cells

A total of 50 electroporations with Molly targeting plasmid were carried out in this experiment, each of which was plated into separate 96 well plates. A total of 10 viable, anti-CD20 antibody secreting clones were 15 obtained and expanded to 120ml spinner flasks. Genomic DNA was isolated from all clones, and Southern analyses were subsequently performed to determine whether the clones represented single homologous recombination events or whether additional random integrations had 20 occurred in the same cells. The methods for DNA isolation and Southern hybridization were as described in the previous section. Genomic DNA was digested with EcoRI and probed with a PCR generated digoxigenin labelled probe to a segment of the CD20 heavy chain constant 25 region. The results of this Southern analysis are presented in figure 6. As can be seen in the figure, 8 of

- 36 -

the 10 clones show a single band hybridizing to the CD20 probe, indicating a single homologous recombination event has occurred in these cells. Two of the ten, clones 24G2 and 28C9, show the presence of additional band(s), indicative of an additional random integration elsewhere in the genome.

We examined the expression levels of anti-CD20 antibody in all ten of these clones, the data for which is shown in Table 1, below.

10

Table 1:

Expression Level of Anti-CD20
Secreting Homologous Integrants

	<u>Clone</u>	<u>Anti-CD20, pg/c/d</u>
	20F4	3.5
15	25E1	2.4
	42F9	1.8
	39G11	1.5
	21C7	1.3
	50G10	0.9
20	29F9	0.8
	5F9	0.3
	-----	-----
	28C9*	4.5
	24G2*	2.1

- 37 -

5

* These clones contained additional randomly integrated copies of anti-CD20. Expression levels of these clones therefore reflect a contribution from both the homologous and random sites.

Expression levels are reported as picogram per cell per day (pg/c/d) secreted by the individual clones, and represented the mean levels obtained from three separate ELISAs on samples taken from 120 mL spinner flasks.

10 As can be seen from the data, there is a variation in antibody secretion of approximately ten fold between the highest and lowest clones. This was somewhat unexpected as we anticipated similar expression levels from all clones due to the fact the anti-CD20 genes are all
15 integrated into the same Desmond marked site. Nevertheless, this observed range in expression extremely small in comparison to that seen using any traditional random integration method or with our translationally impaired vector system.

20 Clone 20F4, the highest producing single copy integrant was selected for further study. Table 2 (below) presents ELISA and cell culture data from seven day production runs of this clone.

- 38 -

Table 2:

7 Day Production Run Data for 20F4

Day	% Viable	viable/ml ($\times 10^5$)	T_{x2}(hr)	mg/L	pg/c/d
1	96	3.4	31	1.3	4.9
5	2	94	6	2.5	3.4
	3	94	9.9	33	4.7
	4	90	17.4	30	6.8
	5	73	14		8.3
	6	17	3.5		9.5

10

Clone 20F4 was seeded at 2×10^5 ml in a 120ml spinner flask on day 0. On the following six days, cell counts were taken, doubling times calculated and 1ml samples of supernatant removed from the flask and analyzed for secreted anti-CD20 by ELISA.

15

This clone is secreting on average, 3-5pg antibody/- cell/day, based on this ELISA data. This is the same level as obtained from other high expressing single copy clones obtained previously in our laboratory using the previously developed translationally impaired random integration vectors. This result indicates the following:

20

- (1) that the site in the CHO genome marked by the Desmond marking vector is highly transcriptionally active, and therefore represents an excellent site from which to express recombinant proteins, and

25

- 39 -

(2) that targeting by means of homologous recombination can be accomplished using the subject vectors and occurs at a frequency high enough to make this system a viable and desirable alternative to random integration
5 methods.

To further demonstrate the efficacy of this system, we have also demonstrated that this site is amplifiable, resulting in even higher levels of gene expression and protein secretion. Amplification was achieved by plating serial dilutions of 20F4 cells, starting at a density of 2.5×10^4 cells/ml, in 96 well tissue culture dishes, and culturing these cells in media (CHO-SSFMII) supplemented with 5, 10, 15 or 20nM methotrexate. Antibody secreting clones were screened using standard ELISA
10 techniques, and the highest producing clones were expanded and further analyzed. A summary of this amplification experiment is presented in Table 3 below.
15

- 40 -

Table 3:

Summary of 20F4 Amplification

nM MTX	# Wells Assayed	Expression Level mg/l 96 well	# Wells Expanded	Expression Level pg/c/d from spinner
10	56	3-13	4	10-15
15	27	2-14	3	15-18
20	17	4-11	1	ND

Methotrexate amplification of 20F4 was set up as described in the text, using the concentrations of methotrexate indicated in the above table. Supernatants from all surviving 96 well colonies were assayed by ELISA, and the range of anti-CD20 expressed by these clones is indicated in column 3. Based on these results, the highest producing clones were expanded to 120ml spinners and several ELISAs conducted on the spinner supernatants to determine the pg/cell/day expression levels, reported in column 5.

The data here clearly demonstrates that this site can be amplified in the presence of methotrexate. Clones from the 10 and 15nM amplifications were found to produce on the order of 15-20pg/cell/day.

A 15nM clone, designated 20F4-15A5, was selected as the highest expressing cell line. This clone originated from a 96 well plate in which only 22 wells grew, and was therefore assumed to have arisen from a single cell.

25 A 15nM clone, designated 20F4-15A5, was selected as the highest expressing cell line. This clone originated

- 41 -

from a 96 well plate in which only 22 wells grew, and was therefore assumed to have arisen from a single cell. The clone was then subjected to a further round of methotrexate amplification. As described above, serial 5 dilutions of the culture were plated into 96 well dishes and cultured in CHO-SS-FMII medium supplemented with 200, 300 or 400nM methotrexate. Surviving clones were screened by ELISA, and several high producing clones were expanded to spinner cultures and further analyzed.

10 A summary of this second amplification experiment is presented in Table 4.

Table 4:
Summary of 20F4-15A5 Amplification

	nM MTX	# Wells Assayed	Expression Level mg/l 96 well	# Wells Expanded	Expression Level pg/c/d, spinner
15	200	67	23-70	1	50-60
	250	86	21-70	4	55-60
	300	81	15-75	3	40-50

20 Methotrexate amplifications of 20F4-15A5 were set up and assayed as described in the text. The highest producing wells, the numbers of which are indicated in column 4, were expanded to 120ml spinner flasks. The expression levels of the cell lines derived from these wells is recorded as pg/c/d in column 5.

25 The highest producing clone came from the 250nM methotrexate amplification. The 250nM clone, 20F4-15A5-250A6 originated from a 96 well plate in which only wells

- 42 -

grew, and therefore is assumed to have arisen from a single cell. Taken together, the data in Tables 3 and 4 strongly indicates that two rounds of methotrexate amplification are sufficient to reach expression levels of 5 60pg/cell/day, which is approaching the maximum secretion capacity of immunoglobulin in mammalian cells (Reff, M.E., *Curr. Opin. Biotech.*, 4:573-576 (1993)). The ability to reach this secretion capacity with just 10 two amplification steps further enhances the utility of this homologous recombination system. Typically, random integration methods require more than two amplification steps to reach this expression level and are generally less reliable in terms of the ease of amplification. Thus, the homologous system offers a more efficient and 15 time saving method of achieving high level gene expression in mammalian cells.

EXAMPLE 5

Expression of Anti-Human CD23 Antibody
in Desmond Marked CHO Cells

20 CD23 is low affinity IgE receptor which mediates binding of IgE to B and T lymphocytes (Sutton, B.J., and Gould, H.J., *Nature*, 366:421-428 (1993)). Anti-human CD23 monoclonal antibody 5E8 is a human gamma-1 monoclonal antibody recently cloned and expressed in our 25 laboratory. This antibody is disclosed in commonly

assigned Serial No. 08/803,085, filed on February 20, 1997.

The heavy and light chain genes of 5E8 were cloned into the mammalian expression vector N5KG1, a derivative 5 of the vector NEOSPLA (Barnett et al, in *Antibody Expression and Engineering*, H.Y Yang and T. Imanaka, eds., pp27-40 (1995)) and two modifications were then made to the genes. We have recently observed somewhat higher secretion of immunoglobulin light chains compared to 10 heavy chains in other expression constructs in the laboratory (Reff et al, 1997, unpublished observations). In an attempt to compensate for this deficit, we altered the 5E8 heavy chain gene by the addition of a stronger promoter/enhancer element immediately upstream of the 15 start site. In subsequent steps, a 2.9kb DNA fragment comprising the 5E8 modified light and heavy chain genes was isolated from the N5KG1 vector and inserted into the targeting vector Mandy. Preparation of 5E8-containing Molly and electroporation into Desmond 15C9 CHO cells 20 was essentially as described in the preceding section.

One modification to the previously described protocol was in the type of culture medium used. Desmond marked CHO cells were cultured in protein-free CD-CHO medium (Gibco-BRL, catalog # AS21206) supplemented with 25 3mg/L recombinant insulin (3mg/mL stock, Gibco-BRL, catalog # AS22057) and 8mM L-glutamine (200mM stock, Gibco-BRL, catalog # 25030-081). Subsequently, trans-

fected cells were selected in the above medium supplemented with 400 μ g/mL geneticin. In this experiment, 20 electroporations were performed and plated into 96 well tissue culture dishes. Cells grew and secreted anti-5 CD23 in a total of 68 wells, all of which were assumed to be clones originating from a single G418 cell. Twelve of these wells were expanded to 120ml spinner flasks for further analysis. We believe the increased number of clones isolated in this experiment (68 compared with 10 for anti-CD20 as described in Example 4) is due to a higher cloning efficiency and survival rate 10 of cells grown in CD-CHO medium compared with CHO-SS-FMII medium. Expression levels for those clones analyzed in spinner culture ranged from 0.5-3pg/c/d, in close agreement with the levels seen for the anti-CD20 15 clones. The highest producing anti-CD23 clone, designated 4H12, was subjected to methotrexate amplification in order to increase its expression levels. This amplification was set up in a manner similar to that described 20 for the anti-CD20 clone in Example 4. Serial dilutions of exponentially growing 4H12 cells were plated into 96 well tissue culture dishes and grown in CD-CHO medium supplemented with 3mg/L insulin, 8mM glutamine and 30, 35 or 40nM methotrexate. A summary of this 25 amplification experiment is presented in Table 5.

Table 5:

- 45 -

Summary of 2H12 Amplification

nM MTX	# Wells Assayed	Expression Level mg/l 96 well	# Wells Expanded	Expression Level pg/c/d from spinner
30	100	6-24	8	10-25
35	64	4-27	2	10-15
5	40	4-20	1	ND

The highest expressing clone obtained was a 30nM clone, isolated from a plate on which 22 wells had grown. This clone, designated 4H12-30G5, was reproducibly secreting 18-22pg antibody per cell per day. This is the same range of expression seen for the first amplification of the anti CD20 clone 20F4 (clone 20F4-15A5 which produced 15-18pg/c/d, as described in Example 4). This data serves to further support the observation that amplification at this marked site in CHO is reproducible and efficient. A second amplification of this 30nM cell line is currently underway. It is anticipated that saturation levels of expression will be achievable for the anti-CD23 antibody in just two amplification steps, as was the case for anti-CD20.

20

EXAMPLE 6

Expression of Immunoadhesin in Desmond Marked CHO Cells

CTLA-4, a member of the Ig superfamily, is found on the surface of T lymphocytes and is thought to play a role in antigen-specific T-cell activation (Dariavach et al, *Eur. J. Immunol.*, 18:1901-1905 (1988); and Linsley et al, *J. Exp. Med.*, 174:561-569 (1991)). In order to further study the precise role of the CTLA-4 molecule in the activation pathway, a soluble fusion protein comprising the extracellular domain of CTLA-4 linked to a truncated form of the human IgG1 constant region was

created (Linsley et al (Id.)). We have recently expressed this CTLA-4 Ig fusion protein in the mammalian expression vector BLECH1, a derivative of the plasmid NEOSPLA (Barnett et al, in Antibody Expression and Engineering, H.Y Yang and T. Imanaka, eds., pp27-40 (1995)). An 800bp fragment encoding the CTLA-4 Ig was isolated from this vector and inserted between the SacII and BglII sites in Molly.

Preparation of CTLA-4Ig-Molly and electroporation into Desmond clone 15C9 CHO cells was performed as described in the previous example relating to anti-CD20. Twenty electroporations were carried out, and plated into 96 well culture dishes as described previously. Eighteen CTLA-4 expressing wells were isolated from the 96 well plates and carried forward to the 120ml spinner stage. Southern analyses on genomic DNA isolated from each of these clones were then carried out to determine how many of the homologous clones contained additional random integrants. Genomic DNA was digested with BglII and probed with a PCR generated digoxigenin labelled probe to the human IgG1 constant region. The results of this analysis indicated that 85% of the CTLA-4 clones are homologous integrants only; the remaining 15% contained one additional random integrant. This result corroborates the findings from the expression of anti-CD20 discussed above, where 80% of the clones were single homologous integrants. Therefore, we can conclude

- 47 -

that this expression system reproducibly yields single targeted homologous integrants in at least 80% of all clones produced.

Expression levels for the homologous CTLA4-Ig
5 clones ranged from 8-12 pg/cell/day. This is somewhat higher than the range reported for anti-CD20 antibody and anti-CD23 antibody clones discussed above. However, we have previously observed that expression of this molecule using the intronic insertion vector system also
10 resulted in significantly higher expression levels than are obtained for immunoglobulins. We are currently unable to provide an explanation for this observation.

EXAMPLE 7

Targeting Anti-CD20 to an alternate Desmond Marked CHO Cell Line

As we described in a preceding section, we obtained
5 single copy Desmond marked CHO cell lines (see Figures
4 and 5). In order to demonstrate that the success of
our targeting strategy is not due to some unique proper-
ty of Desmond clone 15C9 and limited only to this clone,
we introduced anti-CD20 Molly into Desmond clone 9B2
(lane 6 in figure 4, lane 1 in figure 5). Preparation
of Molly DNA and electroporation into Desmond 9B2 was
exactly as described in the previous example pertaining
20 to anti-CD20. We obtained one homologous integrant from
this experiment. This clone was expanded to a 120ml
25

spinner flask, where it produced on average 1.2pg anti-CD20/cell/day. This is considerably lower expression than we observed with Molly targeted into Desmond 15C9. However, this was the anticipated result, based on our 5 northern analysis of the Desmond clones. As can be seen in Figure 5, mRNA levels from clone 9B2 are considerably lower than those from 15C9, indicating the site in this clone is not as transcriptionally active as that in 15C9. Therefore, this experiment not only demonstrates 10 the reproducibility of the system - presumably any marked Desmond site can be targeted with Molly - it also confirms the northern data that the site in Desmond 15C9 is the most transcriptionally active.

From the foregoing, it will be appreciated that, 15 although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without diverting from the scope of the invention. Accordingly, the invention is not limited by the appended claims.

WHAT IS CLAIMED IS:

1. A method for inserting a desired DNA at a target site in the genome of a mammalian cell which comprises the following steps:

5 (i) transfected or transforming a mammalian cell with a first plasmid ("marker plasmid") containing the following sequences:

(a) a region of DNA that is heterologous to the mammalian cell genome which when integrated in the
10 mammalian cell genome provides a unique site for homologous recombination;

(b) a DNA fragment encoding a portion of a first selectable marker protein; and

15 (c) at least one other selectable marker DNA that provides for selection of mammalian cells which have been successfully integrated with the marker plasmid;

(ii) selecting a cell which contain the marker plasmid integrated in its genome;

20 (iii) transfected or transforming said selected cell with a second plasmid ("target plasmid") which contains the following sequences:

(a) a region of DNA that is identical or is sufficiently homologous to the unique region in the
25 marker plasmid such that this region of DNA can recombine with said DNA via homologous recombination;

- 50 -

(b) a DNA fragment encoding a portion of the same selectable marker contained in the marker plasmid, wherein the active selectable marker protein encoded by said DNA is only produced if said fragment is expressed 5 in association with the fragment of said selectable marker DNA contained in the marker plasmid; and

(iv) selecting cells which contain the target plasmid integrated at the target site by screening for the expression of the first selectable marker protein.

10 2. The method of Claim 1, wherein the DNA fragment encoding a fragment of a first selectable marker is an exon of a dominant selectable marker.

15 3. The method of Claim 2, wherein the second plasmid contains the remaining exons of said first selectable marker.

4. The method of Claim 3, wherein at least one DNA encoding a desired protein is inserted between said exons of said first selectable marker contained in the target plasmid.

20 5. The method of Claim 4, wherein a DNA encoding a dominant selectable marker is further inserted between the exons of said first selectable marker contained in

the target plasmid to provide for co-amplification of the DNA encoding the desired protein.

6. The method of Claim 3, wherein the first dominant selectable marker is selected from the group consisting of neomycin phosphotransferase, histidinol dehydrogenase, dihydrofolate reductase, hygromycin phosphotransferase, herpes simplex virus thymidine kinase, adenosine deaminase, glutamine synthetase, and hypoxanthine-guanine phosphoribosyl transferase.

10 7. The method of Claim 4, wherein the desired protein is a mammalian protein.

8. The method of Claim 7, wherein the protein is an immunoglobulin.

15 9. The method of Claim 1, which further comprises determining the RNA levels of the selectable marker (c) contained in the marker plasmid prior to integration of the target vector.

20 10. The method of Claim 9, wherein the other selectable marker contained in the marker plasmid is a dominant selectable marker selected from the group consisting of histidinol dehydrogenase, herpes simplex

thymidine kinase, hydromycin phosphotransferase, adenosine deaminase and glutamine synthetase.

11. The method of Claim 1, wherein the mammalian cell is selected from the group consisting of Chinese hamster ovary (CHO) cells, myeloma cells, baby hamster kidney cells, COS cells, NSO cells, HeLa cells and NIH 3T3 cells.

12. The method of Claim 11, wherein the cell is a CHO cell.

10 13. The method of Claim 1, wherein the marker plasmid contains the third exon of the neomycin phosphotransferase gene and the target plasmid contains the first two exons of the neomycin phosphotransferase gene.

15 14. The method of Claim 1, wherein the marker plasmid further contains a rare restriction endonuclease sequence which is inserted within the region of homology.

20 15. The method of Claim 1, wherein the unique region of DNA that provides for homologous recombination is a bacterial DNA, a viral DNA or a synthetic DNA.

- 53 -

16. The method of Claim 1, wherein the unique region of DNA that provides for homologous recombination is at least 300 nucleotides.

5 17. The method of Claim 16, wherein the unique region of DNA ranges in size from about 300 nucleotides to 20 kilobases.

18. The method of claim 17, wherein the unique region of DNA preferably ranges in size from 2 to 10 kilobases.

10 19. The method of Claim 1, wherein the first selectable marker DNA is split into at least three exons.

20. The method of Claim 1, wherein the unique region of DNA that provides for homologous recombination 15 is a bacterial DNA, an insect DNA, a viral DNA or a synthetic DNA.

21. The method of Claim 20, wherein the unique region of DNA does not contain any functional genes.

20 22. A vector system for inserting a desired DNA at a target site in the genome of a mammalian cell which comprises at least the following:

(i) a first plasmid ("marker plasmid") containing at least the following sequences:

(a) a region of DNA that is heterologous to the mammalian cell genome which when integrated in the mammalian cell genome provides a unique site for homologous recombination;

(b) a DNA fragment encoding a portion of a first selectable marker protein; and

(c) at least one other selectable marker DNA that provides for selection of mammalian cells which have been successfully integrated with the marker plasmid; and

(ii) a second plasmid ("target plasmid") which contains at least the following sequences:

(a) a region of DNA that is identical or is sufficiently homologous to the unique region in the marker plasmid such that this region of DNA can recombine with said DNA via homologous recombination;

(b) a DNA fragment encoding a portion of the same selectable marker contained in the marker plasmid, wherein the active selectable marker protein encoded by said DNA is only produced if said fragment is expressed in association with the fragment of said selectable marker DNA contained in the marker plasmid.

23. The vector system of Claim 22, wherein the DNA fragment encoding a fragment of a first selectable marker is an exon of a dominant selectable marker.

24. The vector system of Claim 23, wherein the
5 second plasmid contains the remaining exons of said
first selectable marker.

25. The vector system of Claim 24, wherein at
least one DNA encoding a desired protein is inserted
between said exons of said first selectable marker con-
10 tained in the target plasmid.

26. The vector system of Claim 24, wherein a DNA
encoding a dominant selectable marker is further insert-
ed between the exons of said first selectable marker
contained in the target plasmid to provide for co-ampli-
15 fication of the DNA encoding the desired protein.

27. The vector system of Claim 24, wherein the
first dominant selectable marker is selected from the
group consisting of neomycin phosphotransferase,
histidinol dehydrogenase, dihydrofolate reductase,
20 hygromycin phosphotransferase, herpes simplex virus
thymidine kinase, adenosine deaminase, glutamine synthe-
tase, and hypoxanthine-guanine phosphoribosyl transfer-
ase.

- 56 -

28. The vector system of Claim 25, wherein the desired protein is a mammalian protein.

29. The vector system of Claim 28, wherein the protein is an immunoglobulin.

5 30. The vector system of Claim 22, wherein the other selectable marker contained in the marker plasmid is a dominant selectable marker selected from the group consisting of histidinol dehydrogenase, herpes simplex thymidine kinase, hydromycin phosphotransferase, adenosine deaminase and glutamine synthetase.

10 31. The vector system of Claim 22, which provides for insertion of a desired DNA at a targeted site in the genome of a mammalian cell selected from the group consisting of Chinese hamster ovary (CHO) cells, myeloma cells, baby hamster kidney cells, COS cells, NSO cells, HeLa cells and NIH 3T3 cells.

15 32. The vector system of Claim 31, wherein the mammalian cell is a CHO cell.

20 33. The vector system of Claim 22, wherein the marker plasmid contains the third exon of the neomycin phosphotransferase gene and the target plasmid contains

- 57 -

the first two exons of the neomycin phosphotransferase gene.

34. The vector system of Claim 22, wherein the marker plasmid further contains a rare restriction endonuclease sequence which is inserted within the region of homology.

35. The vector system of Claim 22, wherein the unique region of DNA that provides for homologous recombination is a bacterial DNA, a viral DNA or a synthetic DNA.

36. The vector system of Claim 22, wherein the unique region of DNA (a) contained in the marker plasmid vector system that provides for homologous recombination is at least 300 nucleotides.

37. The vector system of Claim 36, wherein the unique region of DNA ranges in size from about 300 nucleotides to 20 kilobases.

38. The vector system of Claim 37, wherein the unique region of DNA preferably ranges in size from 2 to 10 kilobases.

- 58 -

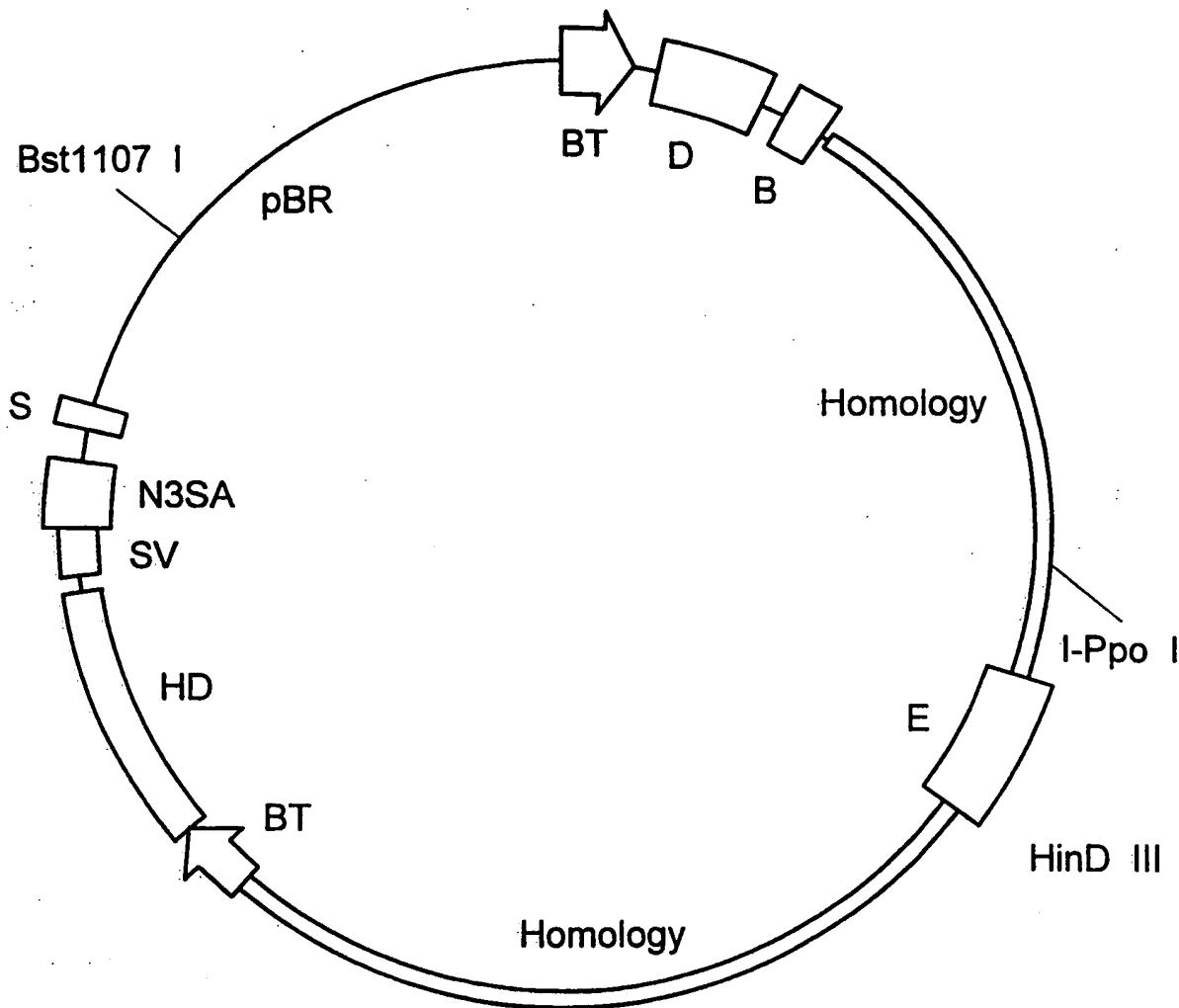
39. The vector system of Claim 22, wherein the first selectable marker DNA is split into at least three exons.

40. The vector system of Claim 22, wherein the
5 unique region of DNA that provides for homologous recombination is a bacterial DNA, an insect DNA, a viral DNA or a synthetic DNA.

41. The vector system of Claim 40, wherein the
10 unique region of DNA does not contain any functional genes.

1/75

FIG. 1A DESMOND



HD = *Salmonella HisD Gene*

N3 = *Neomycin Phosphotransferase Exon 3*

D = *Murine Dihydrofolate reductase*

E = *Cytomegalovirus and SV40 Enhancers*

SA = *Splice acceptor*

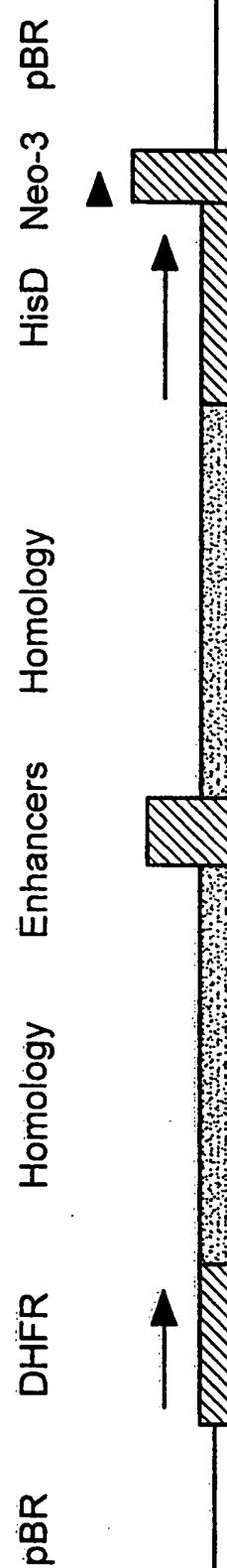
BT = *Mouse Beta Globin Major Promoter*

B = *Bovine Growth Hormone Polyadenylation*

S = *SV40 Early Polyadenylation*

SV = *SV40 Late Polyadenylation*

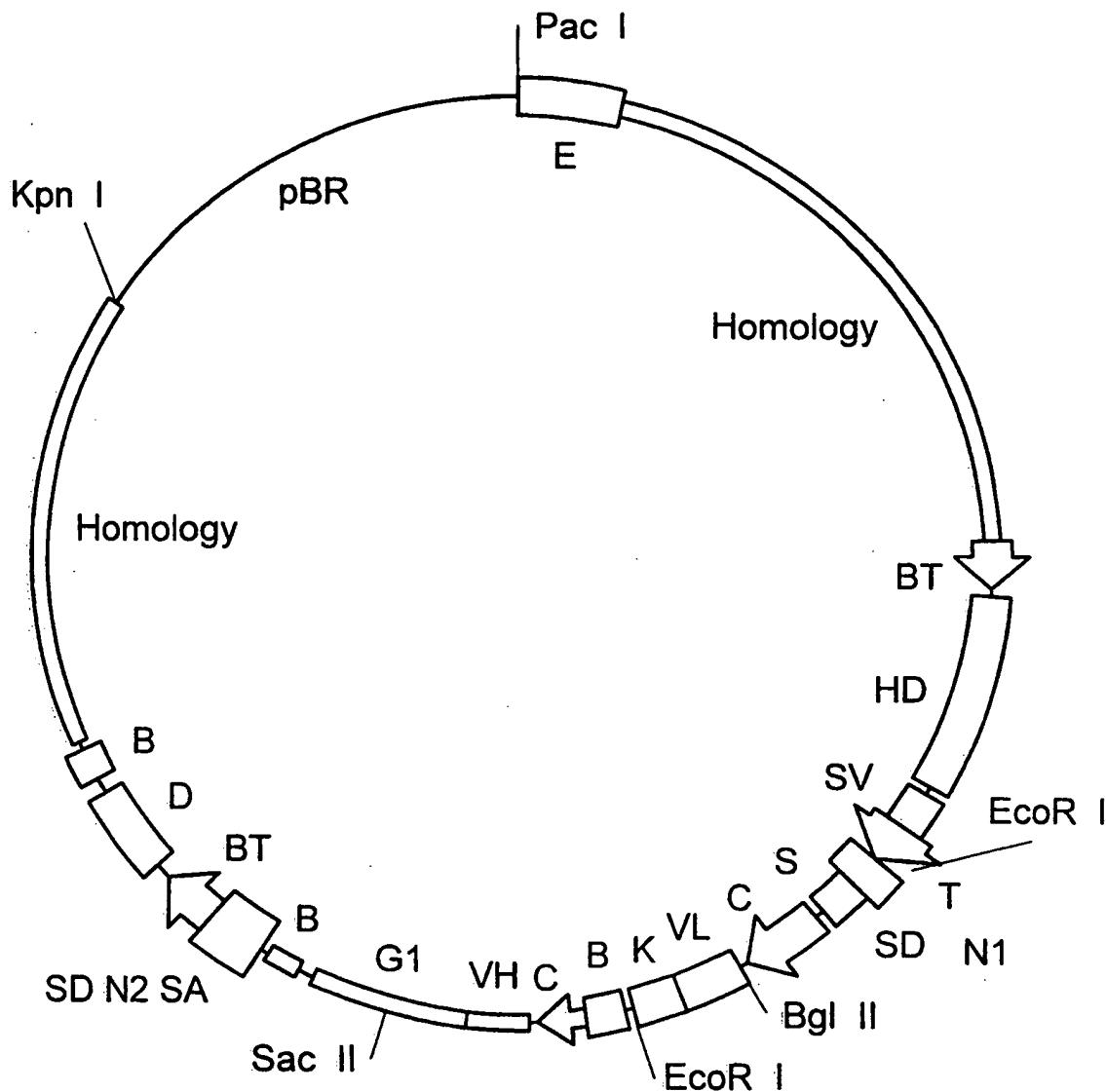
2/75

Desmond**14,683 bp Bst1107 I linear****FIG. 1B**

3/75

Molly

FIG. 2A



D = Dihydrofolate reductase

N1 + Neomycin Phosphotransferase Exon 1

N2 + Neomycin Phosphotransferase Exon 2

VL = Anti-CD20 Light chain leader + Variable

K = Human Kappa Constant

VH = Anti-CD20 Heavy chain Leader + Variable

G1 = Human Gamma 1 Constant

HD = Salmonella Histidinol Dehydrogenase

E = CMV and SV40 enhancers S = SV40 Origin

SD = Splice donor SA = Splice acceptor

C = CMV promoter/enhancer

T = HSV TK promoter and Poloma enhancers

BT = Mouse Beta Globin Major Promoter

SV = SV40 Late Polyadenylation

B = Bovine Growth Hormone Polyadenylation

4/75

Molly
15,987 bp Pac I, Kpn I fragment

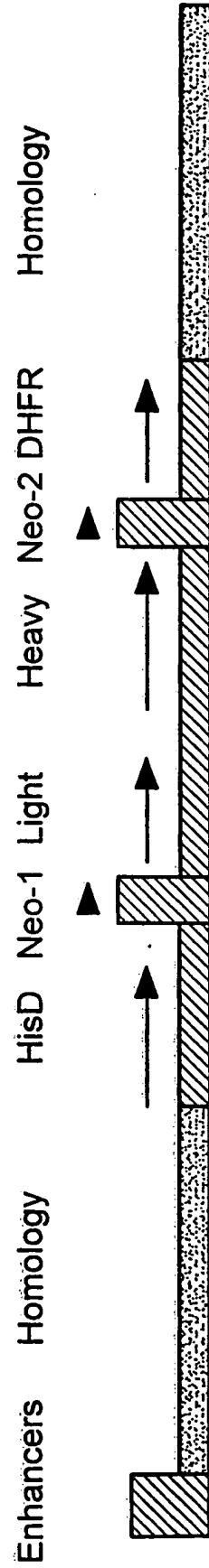
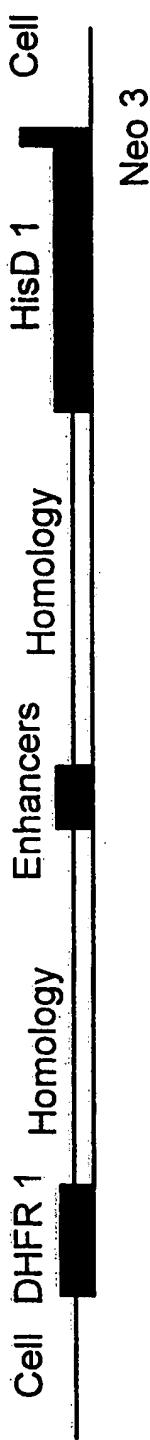


FIG. 2B

FIG. 3

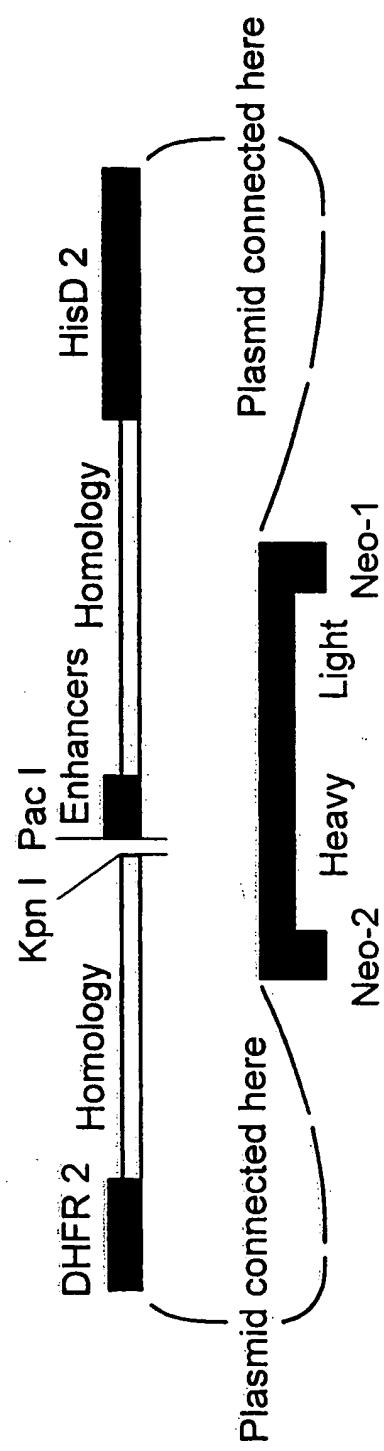
Homologous Recombination

Desmond in CHO

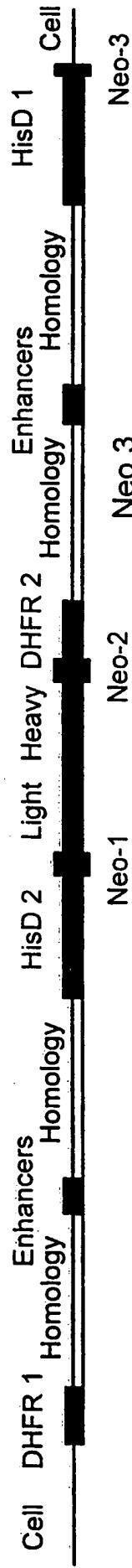


5/75

Molly



Single crossover in CHO



6 / 75

Southern Analysis of Desmond Marked CHO Cells

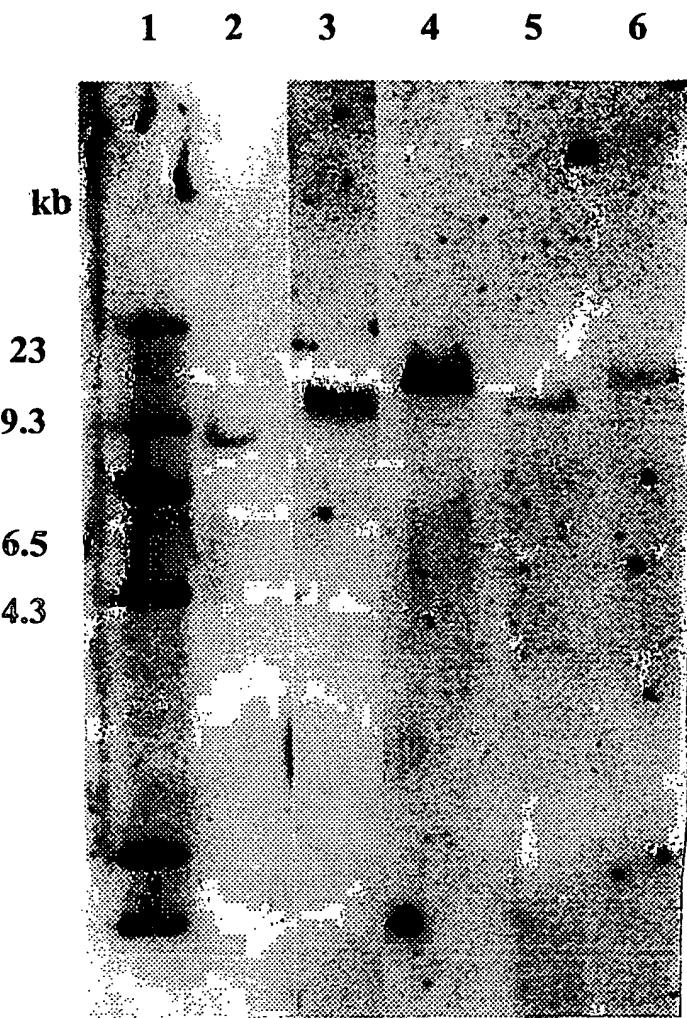


FIG. 4

7/75

Northern Analysis of Desmond Marked CHO Cells

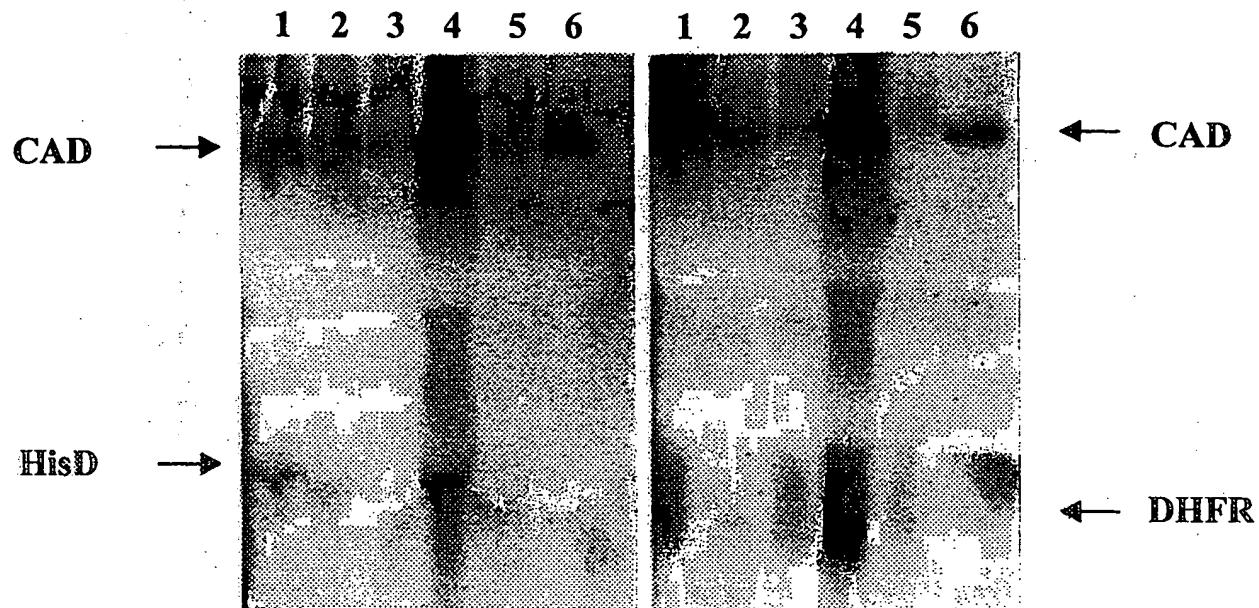


FIG. 5

8 / 75

Southern Analysis of Anti CD20 Integrants in Marked CHO Cells

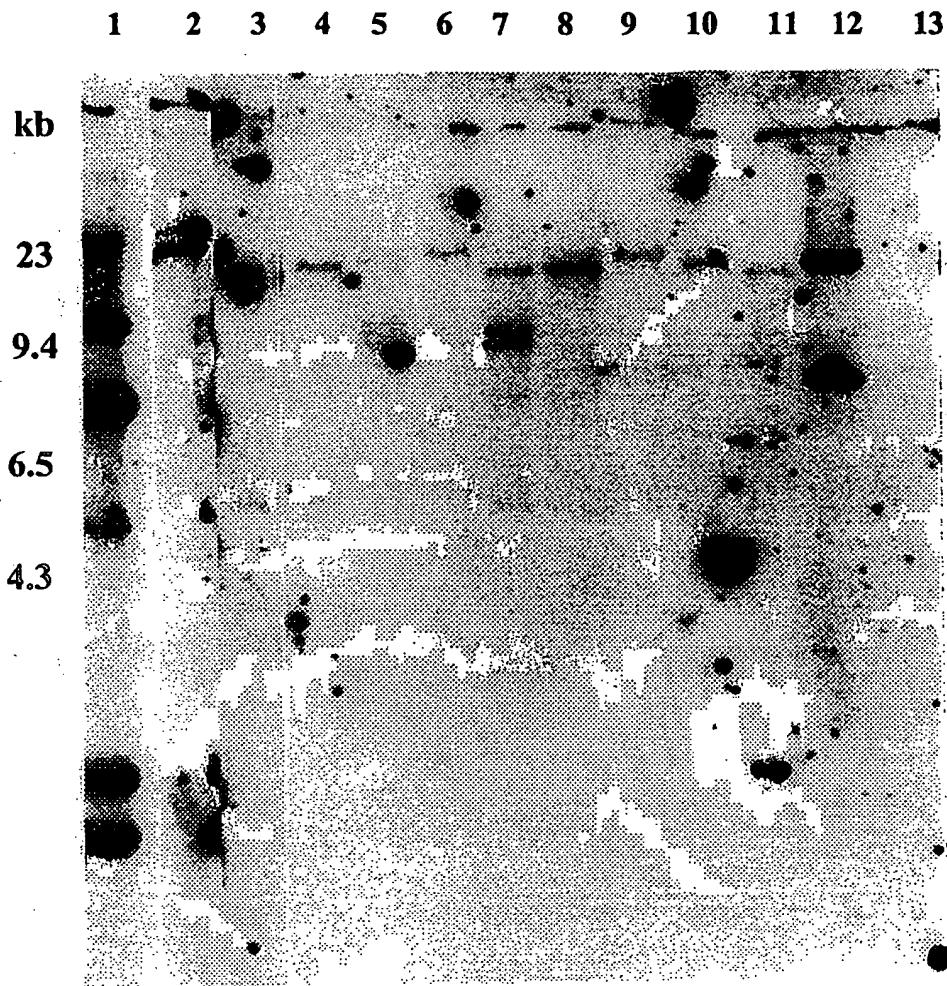


FIG. 6

FIG. 7A

TTCTAGGACC TAGGGCGGCC AGCTAGTAGC TTTGCTTCTC AATTTCCTTAT TTGCATAATG
60
AGAAAAAAAG GAAAATTAAAT TTTAACACCCA ATTCAAGTAGT TGATTGAGCA AATGCGTTG
120
CAAAAGGAT GCTTTAGAGA CAGTGTCTC TGCACAGATA AGGACAAACA TTATTCAAGAG
180
GGAGTACCCA GAGCTGAGAC TCCTAAGCCA GTGAGTGGCA CAGCATTCTA GGGAGAAATA
240
TGCTTGTCA CACCGAAGGCC TGATTCCGTA GAGCCACACC TTGGTAAGGG CCAAATCTGCT
300
CACACAGGAT AGAGAGGGCA GGAGGCCAGGG CAGAGCATAT AAGGGTAGGGT AGGATCAGTT
360
GCTCCTCACA TTTGCTTCTG ACATAGTTGT GTTGGGAGGCT TGATAGCTT GGACAGGCTCA
420
GGGCTGCGAT TTCGGCCAA ACTTGACGGC AATCCTAGCG TGAAGGGCTGG TAGGATTTTA
480
TCCCCGCTGC CATCATGGTT CGACCATGTA ACTGCATCGT CGCCGTTGCC CAAATATGG
540
GGATTGGCAA GAACGGAGAC CTACCCCTGGC CTCCGCTCAG GAACGGAGTTC AAGTACTTCC
600
AAAGAATGAC CACAACCTCT TCAGTGGAG GTAAACAGAA TCTGGTGATT ATGGGTAGGA
660

FIG. 7B

AAACCTGGTT CTCCATTCT GAGAAGAATC GACCTTTAAA GGACAGAAATT AATATAAGTTC
720
TCAGTAGAGA ACTCAAAGAA CCACCACGAG GAGGCTCATT TCTTGCCTAA AGTTGGATG
780
ATGCCCTTAAG ACTTATTGAA CAACCGGAAT TGGCAAGTAA AGTAGACATG GTTTGGATAG
840
TCGGAGGCAG TTCTGTTAC CAGGAAGCCA TGAATCAACC AGGCCACCTT AGACTCTTG
900
TGACAAGGAT CATGCAGGAA TTTGAAAGTG ACACGTTTT CCCAGAAATT GATTGGGA
960
AATATAACT TCTCCAGAA TACCCAGGCG TCCTCTCTGA GGTCCAGGAG GAAAAGGCA
1020
TCAAGTATAA GTTGAAAGTC TAGGAGAAGA AAGACTAACAA GGAAGATGCT TTCAAGTTCT
1080
CTGCTCCCT CCTAAAGCTA TGGCATTTTTA TAAGACCATG GGACCTTTGC TGGCTTTAGA
1140
TCAGCCTCGA CTGTGCCTTC TAGTTGCCAG CCATCTGTTG TTTGCCCTC CCCCCTGCCT
1200
TCCCTGACCC TGGAAAGGTGC CACTCCCCACT GTCCCTTCCT AATAAAATGA GGAAATTGCA
1260
TCGCATTGTC TGAGTAGGTG TCATTCTATT CTGGGGGGTG GGGTGGGGCA GGACAGCAAG
1320

10/75

FIG. 7C

GGGGAGGATT GGGAAAGACAA TAGCAGGCAT GCTGGGGCAT CGGTGGGCTC TATGGAAACCA
1380
GCTGGGGCTC GAAGCGGGCC CCCATTTCGGC TGGTGGTCAG ATGGGGGATG GCGTGGGACGG
1440
CGGGGGGAC CGTCACACTG AGGTTTCCG CCAGACGCCA CTGCTGCCAG GCGCTGATG
1500
GCCCGGGCTC TGACCATGCG GTCGCGTTCG GTTGCAC TAC GCGTACTGTG AGCCAGAGGTT
1560
GCCCGGGCT CTCCGGCTGC GGTAGTTCAAG GCAGTTCAAT CAACTGTTA CCTTGTTGGAG
1620
CGACATCCAG AGGCACCTCA CCGGCTTGCTA GCGGCTTACC ATCCAGGCC ACCATCCAGT
1680
GCAGGAGCTC GTTATCGCTA TGACGGAAACA GGTATTGCT GGTCACCTCG ATGGTTGCC
1740
CGGATAAACG GAACTGGAAA AACTGCTGCT GGTGTTTGC TTCCGGTCAGC GCTGGATGCG
1800
GCGTGGGTC GGCAAAGACC AGACCGTTCA TACAGAACTG GCGATCGTTC GGCGTATCAC
1860
CAAATCACC GCCGTAAGCC GACCACGGGT TGCCGTTTC ATCATATTTA ATCAGCGACT
1920
GATCCACCCA GTCCCCAGACG AAGCCGCCCT GTAAACGGGG ATACTGACGA AACGCCCTGCC
1980

11/75

FIG. 7D

12/75

AGTATTAGC GAAACCGCCA AGACTGTTAC CCATCGCGTG GGCGGTATTTCG CAAAGGATCA
GGGGGGCGT CTCTCCGGGT AGCGAAAGGCC ATTTCGGTGG ATTCGATTTC GGACCATTTC GGACCCAGGCC
2040
GGAAAGGGCTG GTCTTCATCC ACGCGCGCGT ACATCGGGCA AATAATATCG GTGGCCGTGG
2100
TGTGGCTCC GCCGGCTTC TACTGCACCG GGCGGGAAAGG ATCGACAGAT TTGATCCAGC
2160
GATACAGCGC GTCGTGATTAA GCGCCGTGGC CTGATTTCATT CCCCAGCGAC CAGATGATCA
2220
2280
CACTCGGGTG ATTACGATCG CGCTGCACCA TTTCGGTAC GCCTTCGCTC ATCGCCGGTA
2340
GCCAGCGGG ATCATCGGTC AGACGATTCA TTGGCACCAT GCCGGTGGTT TCAATATTGG
2400
CTTCATCCAC CACATACAGG CCTGTAGCGGT CGCACAGCGT GTACCCACAGGC GGATGGTTCG
2460
GATAATGCGA ACAGCGCAGC GCGTTAAAGT TGTTCGTGCTT CATCAGCAGG ATATCCCTGCA
2520
CCATCGTCTG CTCATCCATG ACCTGACCAT GCAGAGGATG ATGCTCGTGA CGGTTAACGC
2580
CTCGAATCAG CAACGGCTTG CCGTTCAGCA GCAGCAGGACCC ATTTCGAATC CGCACCCCTCGC
2640

FIG. 7E

13/75

GGAAACCGAC ATCGCAGGCT TCTGCCCAA TCAGCGTGTGCC GTCGGCGGTG TGCAGGTTCAA
CCACCGCACG ATAGAGATTTC GGGATTTCGG CGCTCCACAG TTTCGGGTTT TCGACCGTTCA 2700
GACGGCAGTGT GACGGCGATCG GCATAACCAC CAGGCTCATC GATAATTTCAC CGGCCGAAAG 2760
GCCGGGTGCC GCTGGCGACC TGCGTTTCAC CCTGCCATAA AGAAACTGTT ACCCGTAGGT 2820
AGTCACCGCAA CTCGCCGCAC ATCTGAACCT CAGCCTCCAG TACAGCGGG CTGAAATCAT 2880
CATTAAGGCG AGTGGCAACA TGGAAATCGC TGATTGTTGT AGTGGGTTA TGCAGGAAACG 2940
AGACGGTCAAG GAAATGCCG CTCATCCGCC ACATATCCTG ATCTTCCAGA TAACTGCCGT 3000
CACTCCACAG CAGCACCCATC ACCGCCGAGGC GGTTTTCTCC GGCGCGTAA AATGCCGCTCA 3060
GGTCAAATTTC AGACGGCAA CGACTGTCCCT GGCTGTAAACC GACCCACGCC CGGTTGCACC 3120
ACAGATGAAA CGCCCGAGTTA ACGCCATCAA AAATAATTTCG CGTCTGGCCT TCTGTAGCC 3180
AGCTTTCATC AACATTAAAT GTGAGCGAGT AACAAACCCGT CGGATTCTCC GTGGGAACAA 3240
3300

FIG. 7F

14/75

ACGGCGGATT GACCGTAATG GGATAGGTTA CGTTGGTGTAA GATGGGGCAGTAACCGT
GCATCTGCCA GTTTGAGGGG ACGACCGACAG TATCGGCCTC AGGAAGATTCG CACTCCAGCC 3360
AGCTTTCCGG CACTGCTTCTC GGTGCCGGAA ACCAGGCAAA GCGCCCATTCG CCATTCAAGGC 3420
TGCGCAACTG TTGGGAAGGG CGATCGGTGC GGCCCTCTTC GCTATTACGC CAGCTGGCGA 3480
AAGCGGGATG TGCTGCAAGG CGATTAAGTT GGTAACGCC AGGGTTTTCC CAGTCACGAC 3540
GTTGTAAAAC GACTTAATCC GTCGAGGGGC TGCTCGAAG CAGACGACCT TCCGTTGTGC 3600
AGCCAGCGGC GCCTGCGCCG GTGCCACAA CCACCAACCTT CTCCCTGTGCC TAACATTCCA GCGCCTCCAC 3660
TCATACGGGC GGCACCCGGG CCACCAACCTT 3720
CACTACCACC ACCATCGATG TCTGAATTGC CGCCCGCTCC ACCAATGCCG ACGGAACCTC 3780
AACCCGCTGC ACCTTTAGAC GACAGACAAC AATTGTTGGA AGCTATTAGA AACGAAAAAA 3840
ATCGCACCTCG TCTCAGACCG GCTCTTAA GGTAGCTCAA ACCAAAAACG GCGCCCCGAAA 3900
3960

FIG. 7G

CCAGTACAAT AGTTGAGGTG CCGACTGTGT TGCCTAAAGA GACATTTGAG CTTAACCGC
CGTCTGCC ACCGCCACCA CCTCCGGCTC CGCCTCCAC GCCAGCCCCC CCTGCCGCCTC
4020
CACCGATGGT AGATTCA TCAGCTCCAC CACCGCCGCC ATTAGTAGAT TTGCCGGTCTG
4080
AAATGTTACC ACCGGCCTGCA CCATCGCTTT CTACACGTGTT GTCTGAATTAA AAATCGGGCA
4140
CAGTTAGATT GAAACCCGCC CAAAAACGCC CGCAATCAGA AATAATTCCA AAAAGCTCAA
4200
CTACAAATTG GATCGGGAC GTGTTAGGCC ACACAAATTAA TAGGCGTCGT GTGGCTATGG
4260
CAAATCGTC TTGGGAAGCA ACTTCTAACG ACGAGGGTTG GGACGGACGAC GATAATCGGC
4320
CTATAAAAGC TAACACGCC GATGTTAAAT ATGTCCAAGC TACTAGTGGT ACCTTAATTAA
4380
AGGGGGAG AATGGGGGA ACTGGGGGA GTTAGGGCG GGATGGGGAGTTAGGGGC
4440
GGGACTATGG TTGCTGACTA ATTGAGATGC ATGCTTTGCA TACTTCTGCC TGCTGGGAG
4500
4560
CCTGGGGACT TTCCACACCT GGTGCTGAC TAATTGAGAT GCATGCTTGC CATACTTCTG
4620

15/75

FIG. 7H

CCTGCTGGG AGCCTGGGA CTTTCCAC CCTAACTGAC ACACATTCCA CAGAATTAAAT
TCCCCTAGTT ATTAATAGTA ATCAATTACG GGGTCATTAG TTTCATAGCCC ATATATGGAG 4680
TTCCGGTTA CATAACTTAC GGTAAATGGC CCCCTGGCT GACCGCTCAA CGACCCCCGC 4740
CCATTGACGT CAATAATGAC GTATGTTCCC ATAGTAACGCC CAATAGGGAC TTTCCATTGA 4800
CGTCAATGGG TGGACTATT ACGGTAAACT GCCCACTTGG CAGTACATCA AGTGTATCAT 4860
ATGCCAAGTA CGCCCCCTAT TGACGTCAT GACGGTAAAT GGCCCCGCCTG GCATTATGCC 4920
CAGTACATGA CCTTATGGGA CTTTCCTACT TGGCAGTACA TCTACGTATT AGTCATCGCT 4980
ATTACCATGG TGATGGGTT TTGGCAGTAC ATCAATGGGC GTGGATAGGC GTTTGACTCA 5040
CGGGGATTTC CAAGTCTCCA CCCCATTGAC GTCAATGGGA GTTTGTTTG AAGCTTGGCC 5100
GCCCATATAA ACGGCGGCCA GCTTTATTTA ACGTGTAC GTCGAGTCAA TTGTACACTA 5160
ACGACAGTGA TGAAAGAAAT ACAAAAGCGC ATAATATTT GAACGGACGTC GAACCTTTAT 5220
5280

16/75

FIG. 71

TACAAACAA AACACAAACG AATATCGACA AAGCTAGATT GCTGCTACAA GATTTGGCAA
5340
GTTTGTGGC GTTGAGGGAA AATCCATTAG ATAGTCCAGC CATCGGGTCG GAAAAACAAAC
5400
CCTTGTGTGA AACTAATCGA AACCTATTT ACAAAATCTAT TGAGGATT TA ATATTTAAAT
5460
TCAGATATA AGACGGCTGAA AATCATTGTA TTTTCGCTCT AACATACCAC CCTAAAGATT
5520
ATAAATTAA TGAATTATA AAATACATCA GCAACTATAT ATTGATAGAC ATTCCAGTT
5580
TGTGATATA GTTTGTGGT CTCATTACAA TGGCTGTTAT TTTTAACAAAC AACAAACTGC
5640
TCGCAGACAA TAGTATGAA AAGGGAGGGTG AACTGTTTT GTTAAACGGT TCGTACAAACA
5700
TTTTGGAAAG TTATGTTAAT CCGGTGCTGC TAAAAAATGG TGTAATTGAA CTAGAAGGAAG
5760
CTGCGTACTA TGCCGGCAC ATATTGTACA AAACCGACGA TCCCAAATTC ATTGATTATA
5820
TAAATTAAAT ATTAAAGCA ACACACTCCG AAGAACTACC AGAAAATAGC ACTGTTGTAA
5880
ATTACAGAAA AACTATGCC AGCGGGTACTA TACACCCCAT TAAAAAAGAC ATATATTTT
5940

FIG. 7J

18/75

ATGACAACAA AAAATTACT CTATACGATA GATACATATA TGGATAACGAT ATAACATATG
TTAATTTTTA TGAGGAGAAA AATGAAAAAG AGAAGGAATA CGAACGAAAGAA GACCGACAAGG
CGTCTAGTTT ATGTGAAAAT AAAATTATAT TGTGCAAAT TAACTGTGAA TCATTGTGAA
ATGATTAA ATATTACCTC AGCGATTATA ACTACGCCGT TTCAATTATA GATAACACTA
CAAATGTTCT TGTTGCGTTT GGTTTGTATC GTTAATAAAA AACAAATTAA GCATTATAA
TTGTTTATT ATTCAAATAAT TACAAATAGG ATTGAGACCC TTGCAGTTGC CAGCAAACGG
ACAGAGCTTG TCGAGGGAGAG TTGTTGATTG ATTGTTGCTGCC TCCCTGCTGC GTTTTGAC
CGAAGTTCAT GCCAGTCCAG CGTTTTGCA GCAGAAAAGC CGCCGACTTC GGTTTGCGGT
CGCGAGTGAA GATCCCTTTC TTGTTACCGC CAACGCGCAA TATGCCCTGC GAGGTCGCAA
AATCGGGCAA ATTCCATACC TGTTCACCGA CGACGGCGCT GACGCGATCA AAGACGGGT
GATACATATC CAGCCATGCA CACTGATACT CTTCACTCCA CATGTCGGTG TACATTGAGT
6000 6060 6120 6180 6240 6300 6360 6420 6480 6540 6600

FIG. 7K

GCAGCCCCGC TAACGTATCC ACGCCCGTATT CGGTGATGAT AATCGGGCTGA TGCAGTTCT
CCTGCCAGGC CAGAAGTTCT TTTCCAGTA CCTTCTTGC CGTTCCAAA TCGCCGCTTT 6660
GGACATACCA TCCGTAATAA CGGTCAGGGC ACAGGCACATC AAAGAGATCG CTGATGGTAT 6720
CGGTGTGAGC GTCGCAGAAC ATTACATTGA CGCAGGTGAT CGGACGGC GTGC GGGTCCGAGTT 6780
TACGGCGTTGC TTCCGCCAGT GGCGCGAAAT ATTCCCGTGC ACCTTGCGGA CGGGTATCCG 6840
GTTCGTTGGC AATACTCCAC ATCACCACCGC TTGGGTGGTT TTTGTCACGC GCTATCAGCT 6900
CTTAATCGC CTGTAAGTGC GCTTGGTGAG TTTCCCCGTT GACTGCCCTCT TCGTTGTACA 6960
GTTCTTCCG CTTGTTGCC GCTTCGAAAC CAATGCCCTAA AGAGAGGTAA AAGGCCGACAG 7020
CAGCAGTTTC ATCAATCACC ACGATGCCAT GTTCATCTGC CCAGTCGAGC ATCTCTTCAG 7080
CGTAAGGGTA ATGCGAGGTA CGGTAGGGAGT TGGCCCTAAT CCAGTCCATT AATGCGTGGT 7140
CGTGCACCAT CAGCACCGTTA TCGAATCCCTT TGCCACGCAA GTCCGCATCT TCATGACGAC 7200
7260

19/75

FIG. 7L

CAAGCCAGT AAAGTAGAAC GGTTTGTGGT TAATCAGGAA CTGTTGCCCTTCACTGCCA
CTGACCGGAT GCCGACGGCA AGCGGGTAGA TATCACACTC TGTCTGGCTT TTGGCTGTGA
7320
CGCACAGTTC ATAGAGATAA CCTTCACCCG GTTGCAGAG GTGCCAGATTC ACCACTTGCA
7380
AAGTCCCGCT AGTGCCTTGT CCAGTTGCAA CCACCTGTTG ATCCGCATCA CGCAGTTCAA
7440
CGCTGACATC ACCATTGCC ACCACCTGCC AGTCAACAGA CGCGTGGTTA CAGTCTTGC
7500
CGACATGCGT CACTACGGTG ATATCGTCCA CCCAGGTGTT CGGCAGGTGG TAGAGCATT
7560
CGCTGGATG GATTCCGGCA TAGTTAAAGA AATCATGGAA GTAAGATTGC TTTTCTTGC
7620
CGTTTCTGTT GGTAATCACC ATTCCCCGGCG GGATAGTCTG CCAGTTCAGT TCAGTTGTTCA
7680
7740
CACAAACGGT GATACCCCTC GACGGATTAA AGACTTCAG CGGTCAACTA TGAAGAAGTG
7800
TTCGTCTCG TCCCAGTAAG CTATGTCCT AGAATGTAGC CATCCATCCT TGTCATCAA
7860
GGCGTTGGTC GCTTCCGGAT TGTTTACATA ACCGGACATA ATCATAGGTC CTCTGACACA
7920

20/75

FIG. 7M

TAATACGCCT CTCTGATTAA CGCCCCAGCGT TTTCCCCGGTA TCCAGATCCA AACCTTCGC
7980
TTCAAAAAT GGAACAACTT TACCGACCGC GCCGGTTA TCATCCCCCT CGGGTGTAAAT
8040
CAGAATAGCT GATGTAGTCT CAGTGAGCCC ATATCCTTGT CGTATCCCTG GAAGATGGAA
8100
GCGTTTGC A ACCGCTTCCC CGACTTCTTT CGAAAGAGGT GCGCCCCAG AAGCAATTTC
8160
GTGTAAATT A GATAAATCGT ATTTGTCAAT CAGAGTGCCT TTGGCGAAGA ATGAAAATAG
8220
GGTTGGTACT AGCAACGCAC TTTGAATTGT GTAAATCCTGTA AGGGATCGTA AAAACAGCTC
8280
TTCTTCAAAT CTATACATTA AGACGACTCG AAATCTACAT ATCAAATATC CGAGTGTAGT
8340
AACATTCCA AAACCGTGTGAT GGAATGGAAC AACACTAAA ATCGCAGTAT CGGAATGAT
8400
TTGATTGCCA AAAATAGGAT CTCTGGCATG CGAGAATCTA GCGCAGGCAG TTCTATGCGG
8460
AAGGGCCACA CCCTTAGGTA ACCCAGTAGA TCCAGAGGAA TTGTTTTGTC ACGATCAAAG
8520
GACTCTGGTA CAAAATCGTA TTCATTAAA CCGGGAGGTA GATGAGATGT GACGGAAGGTG
8580

FIG. 7N

22/75

TACATCGACT GAAATCCCTG GTAATCCGT TTAGAATTCCA TGATAATAAT TTTCTGGATT
ATTGGTAATT TTTTGCAC GTTCAAAATT TTTGCCAACC CCTTTTGGGA AACAAACACT 8640
ACGGTAGGCT GCGAAATGTT CATACTGTT AGCAATTCACTAC GTTCATTATA AATGTCGTTT 8700
GGGGCGCAA CTGCAACTCC GATAAATAAC GCCCCAAACA CGGGCATAAA GAATTGAAGA 8760
GAGTTTCAC TGCATACGAC GATTCTGTGA TTTGTATTCA GCCCATATCG TTTCATAGCT 8820
TCTGCCAACC GAACGGACAT TTCGAAGTAT TCCGGCTACG TGATGTTCAC CTCGATATGT 8880
GCATCTGTAA AAGGAATTGT TCCAGGAACC AGGGCGTATC TCTTCATAGC CTTATGCAGT 8940
TGCTCTCAG CGGTTCCATT CTCTAGCTTT GCCTCTCAAT TTCTTATTG CATAATGAGA 9000
AAAAAGGAA AATTAATTAA AACACCAATT CAGTAGTTGA TTGAGCAAAT GCGTTGCCAA 9060
AAAGGATGCT TTAGAGACAG TGTTCTCTGC ACAGATAAGG ACAAACATCA TTCAAGGGGA 9120
GTACCCAGAG CTGAGACTCC TAAGCCAGTG AGTGGCACAG CATTCTAGGG AGAAATATGC 9180
9240

FIG. 7P

TTGTCATCAC CGAAGCCTGA TTCCGTAGAG CCACACCTTG GTAAGGGCCA ATCTGCTCAC
ACAGGATAGA GAGGGCAGGA GCCAGGGCAG AGCATATAAG GTGAGGTAGG ATCAGTTGCT
CCTCACATT GCTTCTGACA TAGTTGTGTT GGGAGCTTGG ATCGATCCAC CATGGGCTTC
AATACCCCTGA TTGACTGGAA CAGCTGTAGC CCTGAAACAGC AGCGTGCCT GCTGACGGGT
CCGGCGATT CCGCCTCTGA CAGTATTACC CGGACGGTCA GCGATATTCT GGATAATGCA
AAAACGGCG GTGACGATGC CCTGCGTGA TACAGGGCTA AATTGATAA AACAGAACGTC
ACAGCGCTAC GCGTCACCCC TGAAGAGATC GCCGCCGCCG GCGGCCGTCT GAGCGACGAA
TTAAAACAGG CGATGACCC TCCCCCAA AATATTGAAA CGTTCATTG CGCCAGGACG
CTACCGCTTG TAGATGTGGA AACCCAGCCA GGCGTGCCTT GCCAGCAGGT TACGGTCCC
GTCTCGTCTG TCGGTCTGTA TATTCCGGC GGCTCGGCTC CGCTCTTCTC AACGGTGTG
ATGCTGGCGA CGCCGGCGC CATTGGGGGA CATTAGAAGG TGCTAGAAGG TGCTTCTGTG CTCGCCGCCG
9300
9360
9420
9480
9540
9600
9660
9720
9780
9840
9900
23/75

FIG. 7Q

24/75

CCCATGGCTG ATGAAATCCT CTATGCCGG CAACTGTGT GCGTGCAGGA ATTCTTTAAC
CTCGGGGGCG CGCAGGGAT TGCCGGCTTG GCCTTGGCA GCGAGTCCGT ACCGAAAGTG
9960
GATAAAATT TTGGCCCCGG CAACGCCCTT GTAACCGAAG CCAAACGTCA GGTCA GCCAG
10020
CGTCTGGACG GCGGGCTAT CGATATGCCA GCCGAGCCGT CTGAAGTACT GGTGATCGCA
10080
GACAGGGCG CAACACCGGA TTTCGTCGCT TCTGACCTGC TCTCCCAGAC TGAGCACGGC
10140
10200
CCGGATTCCC AGGTGATCCT GCTGACGCC GATGCTGACA TTGCCCGCAA GGTGGCGGAG
10260
GGGGTAGAAC GTCAACTGGC GGAACTGCC CGGGGGACA CGGCCCTGGCA GCCCCTGAGC
10320
GCCAGTCGTC TGATTGTGAC CAAAGATTAA GCGCAGTGC GCG TCGCCATCTC TAATCAGTAT
GGCCGGAAC ACTTAATCAT CCAGACGCC AATGCGCGC ATTTGGTGA TGCGATTACC
10380
AGCGCAGGCT CGGTATTTCT CGGCGACTGG TCGCCGGAAAT CGGCCGGTGA TTACGCTTCC
10440
GGAACCAACC ATGTTTACCG GACCTATGGC CATACTGCTA CCTGTTCCAG CCTTGGGTTA
10500
10560

FIG. 7R

GGGGATTCC AGAACGGAT GACCGTTTCAG GAACTGTCA
GCATCAACCA TTGAAACATT GGCGGGCA GAACGTTCTGA CCGCCCCATAA AAATGCCGTC
ACCCTGCAG TAAACGCCCT CAAGGAGCAA GCATGAGCAC TGAAAACACT CTCAGCGTC
CTGACTTAGC CCGTGAAAT GTCCGCAACC TGAGAGATCCA GACATGATAA GATACATTGA
TGAGTTGGA CAAACCACAA CTAGAATGCA GTGAAAAAAA TGCTTTATT TGAAATTGG
TGATGCTATT GCTTTATTG TAACCATTAT AAGCTGCAAT AAACAAAGTTA ACAACAAACAA
TTCATTCTCAT TTTATGTTTC AGGTTCAAGGG GGAGGGTGTGG GAGGGTTTTT AAAAGCAAGTA
AAACCTCTAC AAATGTGGTA TGGCTGATT TGATCTCTAG CTCGACGGGG CGCCTGGCCG
CTACTAACTC TCTCCTCCCT CCTTTTCCT GCAGGCTCAA GGCGCGCATG CCCGACGGCG
AGGATCTCGT CGTGACCCAT GGCGATGCCT GCTTGCCAA TATCATGGTG GAAAATGGCC
GCTTTCTGG ATTCAATCGAC TGTGGCCGGC TGGGTGTGGC GGACCGCTAT CAGGACATAG
11220

25/75

FIG. 7S

26/75

CGTTGGCTAC CGGTGATATT GCTGAAGAGGC TTGGGGCGA ATGGGCTGAC CGCTTCCTCG
TGCTTTACGG TATCGCCGCT CCCGATTTCGGC AGCGCATTCGC CTTCTATCGC CTTCTTGCACG
11280
11340
AGTTCTTCTG AGCGGGACTC TGGGGTTCGA AATGACCGAC CAAGCGACGC CCAACCTGCC
11400
ATCACGAGAT TTCGATTCCA CGGCCGCCCTT CTATGAAAGG TTGGGCTTCG GAATCGTTT
11460
11520
CCGGGACGCC GGCTGGATGA TCCTCCAGCG CGGGGATCTC ATGCTGGAGT TCTTCGCCCA
11580
CCCCAACTTG TTTATTGCAG CTTATAATGG TTACAAATAA AGCAATAGCA TCACAAATT
CACAAATAA GCATTTTTT CACTGCATTCT TAGTTGTGGT TTGTCCAAAC TCATCAAATCT
ATCTTATCAT GTCTGGATCG CGGCCCCGTCT CTCTCTAGGCC CTAGGTCTAG ACTTTGGCAGA
11640
11700
ACATATCCAT CGCGTCCGCC ATCTCCAGCA GCCGCACGCC GCGCATCTCG GGAGGCGTTG
GGTCCTGCC ACGGGTGCAC ATGATCGTGC TCCTGTGCTT GAGGACCCGG CTAGGCTGGC
11760
11820
GGGTTGCC TACTGGTTAG CAGAATGAAT CACCGATAACG CGAGCGAACG TGAAGCGACT
11880

FIG. 7T

27/75

GCTGCTGCAA AACGTCTGCG ACCTGAGCAA CAAACATGAAT GGTCTTCGGT TTCCGGTTT
11940
CGTAAAGTCT GGAAACGGGG AAGTCAGGCC CCTGCACCAT TATGTTCCGG ATCTGCATCG
12000
CAGGATGCTG CTGGCTACCC TGTGGAACAC CTACATCTGT ATTAAACGAAG CGCTGGCATT
12060
GACCCTGAGT GATTTTCTC TGGTCCCCGCC GCATCCATAAC CGCCAGTTGT TTACCCCTCAC
12120
AACGTTCCAG TAACCGGGCA TGTTCATCAT CAGTAACCCG TATCGTGAGC ATCCCTCTC
12180
GTTTCATCGG TATCATTACC CCCATGAACA GAAATCCCCC TTACACGGAG GCATCAGTGA
12240
CCAAACAGGA AAAAACGCC CTTAACATGG CCCGCTTTAT CAGAACGCCAG ACATTAACGC
12300
TTCTGGAGAA ACTCAAACGAG CTGGACGGCG ATGAACAGGC AGACATCTGT GAATCGCTTC
12360
ACGACCACGC TGATGAGCTT TACCGCAGCT GCCTCGCGCG TTTGGGTGAT GACGGTGAAA
12420
ACCTCTGACA CATGCAGCTC CCGGAGACGG TCACAGCTTG TCTGTAAGCG GATGCCGGGA
12480
GCAGACAAGC CCGTCAGGGC GCGTCAGCGG GTGTTGGCGG GCAGGCCATGA
12540

FIG. 7U

28/75

CCCAGTCACG TAGCGATAGC GGAGGTGTATA CTGGCTTAAC TATGCCCAT CAGAGCAGAT
12600
TGTACTGAGA GTGCACCATA TGCGGTGTGA AATACCGCAC AGATGGTAA GGAGAAAATA
12660
CCGCATCAGG CGCTCTCCG CTTCCCTCGCT CACTGACTCG CTGGCGCTCGG TCGGTTCGGCT
12720
GGGGCAGCG GTATCAGCTC ACTCAAAGGC GGTAAATACGG TTATCCACAG AATCAGGGGA
12780
TAACGCAGGA AAGAACATGT GAGCAAAAGG CCAGCAAAAG GCCAGGAACC GTAAAAAGGC
12840
CGCGTTGCTG GCGTTTTCC ATAGGCTCCG CCCCCCTGAC GAGGCATCACA AAAATCGACG
12900
CTCAAGTCAG AGGTGGGAA ACCCGACAGG ACTATAAAGA TACCAAGGGT TTCCCCCTGG
12960
AAGCTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC TGTCCGCCCT
13020
TCTCCCTCG GGAAAGCGTGG CGCTTTCTCA TAGCTCACGC TGTAGGTATC TCAGTTGGT
13080
GTAGGTCGTT CGCTCCAAGC TGGGCTGTGT GCACGAACCC CCCGTTCAAGC CCGACCGCTG
13140
CGCCTTATCC GGTAACTATC GTCTTGAGTC CAACCCGGTA AGACACGACT TATCGCCACT
13200

FIG. 7V

29/75

GGCAGCAGGCC ACTGGTAACA GGATTAGCAG AGCGGAGGTAT GTAGGGCGGTG CTACAGAGTT
CTTGAAGTGG TGGCCTAACT ACGGCCTACAC TAGAAGGACA GTATTGGTA TCTGCCGCTCT
GCTGAAGCCA GTTACCTTCG GAAAAAGAGT TGGTAGCTCT TGATCCGGCA AACAAACCAC
CGCTGGTAGC GGTGGTTTTT TTGTTTGCAA GCAGGCAGATT ACGCCGCAGAA AAAAGGATC
TCAAGAAGAT CCTTGATCT TTCTACGGG GTCTGACGCT CAGTGGAACG AAAACTCACG
TTAAGGGATT TTGGTCATGA GATTATCAA AAGGATCTTC ACCTAGATCC TTTTAAATTA
AAAATGAAGT TTAAATCAA TCTAAAGTAT ATATGAGTAA ACTTGGTCTG ACAGTTACCA
ATGCTTAATC AGTGAGGCAC CTATCTCAGC GATCTGTCTA TTTCGTTCAT CCATAGTTGC
CTGACTCCCC GTCGTGTAGA TAACTACGAT ACGGGAGGGC TTACCATCTG GCCCCAGTGC
TGCAATGATA CCGCGAGACC CACGCTCACC GGCTCCAGAT TTATCAGCAA TAAACCAGCC
AGCCGGAAAGG GCCGAGGCC GAAGTGGTCC TGCAACTTTA TCCGGCTCCA TCCAGTCTAT
13260
13320
13380
13440
13500
13560
13620
13680
13740
13800
13860

FIG. 7W

30/75

TAATTGTGTC CGGGAAAGCTA GAGTAAGTAG TTCGCCAGTT AATAGTTTGC GCAACGTTGT
 13920
 TGCATGGCT GCAGGCATCG TGGTGTCAAGG CTCGGTGGTTT GGTATGGCTT CATTCAAGCTC
 13980
 CGGTTCCCAA CGATCAAGGC GAGTTACATG ATCCCCATG TTG TGCAAAA AAGCGGGTTAG
 14040
 CCTCTCGGT CCTCCGATCG TTGTCAGAAG TAAGTTGCC GCAGTGTTAT CACTCATGGT
 14100
 TATGGCAGCA CTGCATAATT CTCTTACTGT CATGCCATCC GTAAGATGCT TTTCTGTGAC
 14160
 TGGTGAGTAC TCAACCAAGT CATTCTGAGA ATAGTGTATG CGGCCGACCGA GTTGCTCTTG
 14220
 CCCGGCGTCA ACACGGATA ATACCGCGCC ACATAGCAGA ACTTTAAAG TGCTCATCAT
 14280
 TGGAACACGT TCTTCGGGC GAAAACCTCTC AAGGATCTTA CCGCTGTGTA GATCCAGTTTC
 14340
 GATGTAACCC ACTCGTGCAC CCAAATGATC TTCAGCATCT TTTACTTTCA CCAGCGTTTC
 14400
 TGGGTGAGCA AAAACAGGAA GGCAAAATGC CGCAAAAGGG GGAATAAGGG CGACACGGAA
 14460
 ATGTTGAATA CTCATACTCT CCCTTTTCA ATATTGA AGCATTATC AGGGTTATTG
 14520

FIG. 7X

TCTCATGAGC GGATACATAT TTGAATGTAT TTAGAAAAAT AAACAAATAG GGGTTCCGCG
14580
CACATTTCCC CGAAAAGTGC CACCTGACGT CTAAGAAACC ATTATTATCA TGACATTAAAC
14640
CTATAAAAT AGGCGTATCA CGAGGCCCTT TCGTCTTCAA GAA
14683

FIG. 8A

TTAAAGG GGCGGGAAT GGCGGGACT GGGCGGGTT AGGGCGGGAA TGGGGGGAGT 60
TAGGGGG ACTATGGTTG CTGACTAATT GAGATGCATG CTTTGCATAC TTCTGCCCTGC 120
TGGGAGCCT GGGGACTTTG CACACCTGGT TGCTGACTAA TTGAGATGCA TGCTTTGCAT 180
ACTTCTGCCT GCTGGGAGC CTGGGAGCT TCCACACCCCT AACTGACACA CATTCACAG 240
AATTAATTCC CCTAGTTATT AATAGTAATC ATTACGGGG TCATTAGGTC ATAGCCCCATA 300
TATGGAGTTCC CGCGTTACAT AACCTTACGGT AAATGGCCCG CCTGGCTGAC GCCAACGA 360 32/75
CCCCCGCCA TTGACCGTCAA TAATGACGTA TGTTCCCATA GTAACGCCAA TAGGGACTTT 420
CCATTGACGT CAATGGGTGG ACTATTACG GTAAACTGCCC CACCTGGCAG TACATCAAGT 480
GTATCATATG CCAAGTACGC CCCCTATTGA CGTCAATGAC GGAAATGGC CGCCTGGCA 540
TTATGCCAG TACATGACCT TATGGGACTT CCCTACTTGG CAGTACATCT ACGTATTAGT 600
CATCGCTATT ACCATGGTGA TGCGGTTTG GCAGTACATC AATGGGGCTG GATAGGGTT 660
TGACTCACCGG GGATTCCCAA GTCTCCACCC CATTGACGTC AATGGGAGTT TGTTTGAAAG 720
CTGGCCGGC CATATAACG GCGGCCAGCT TTATTAAACG TGTTTACGTC GAGTCAAATTG 780
TACACTAACG ACAGTGATGA AAGAAATACA AAAGGCCATA ATATTGGAA CGACGTCGAA 840

FIG. 8B

CCTTTATTAC AAAACAAAC ACAAACGAAT ATCGACAAAG CTAGATTGCT GCTACAAGAT 900
 TTGGCAAGTT TTGTGGCGTT GAGCGAAAAT CCATTAGATA GTCCAGCCAT CGGTTGGAA 960
 AAACAACCT TGTGAAAC TAATCGAAAC CTATTTCACA AATCTATTGA GGATTAAATA 1020
 TTTAAATTCA GATAAAAGA CGCTGAAAAT CATTGATT TCGCTCTAAC ATACCACCCCT 1080
 AAAGATTATA ATTAAATGA ATTATTAAGA TACATCAGCA ACTATATATT GATAGACATT 1140
 TCCAGTTTGT GATATTAGTT TGTGCGTCTC ATTACAATGG CTGTTATTCT TAACAAACAA 1200
 CAACTGCTCG CAGACAATAG TATAGAAAAG GGAGGGTGAAC TGTTTTTGTT TAACGGTTCG 1260
 TACAACATT TGGAAAGTTA TGTAAATCCG GTGCTGCTAA AAAATGGTGT ATTGAACTA 1320
 GAAGAGCTG CGTACTATGC CGGCAACATA TTGTACAAAA CCGACGATCC CAAATTCACT 1380
 GATTATATA ATTAAATAAT TAAAGCAACA CACTCCGAAG AACTACCAGA AAATAGCACT 1440
 GTTGTAAATT ACAGAAAAC TATGCGCAGC GGTACTATAC ACCCCATTAA AAAAGACATA 1500
 TATTTTATG ACAACAAAAA ATTTACTCTA TACGATAGAT ACATATATGG ATACGATAAT 1560
 AACATTGTTA ATTITATGA GGAGAAAAT GAAAAAGAGA AGGAATAACGA AGAAGAAGAC 1620
 GACAAGGGCGT CTAGTTTATG TGAAAATAAA ATTATATTGT CGCAAATTAA CTGTGAATCA 1680

33/75

FIG. 8C

TTGAAATG ATTTAAATA TTACCTCAGC GATTAAACT ACGCGTTTC AATTATAGAT 1740
AATACTACAA ATGTTCTTGT TGCCTGGT TTGTATCGTT AAAAAAAC AAATTTAGCA 1800
TTTATAATTG TTTTATTATT CAATAATTAC AAATAGGATT GAGACCCTTG CAGTGGCCAG 1860
CAAACGGACA GAGCTTGTAG AGGAGAGTTG TTGATTCAATT GTTGCCTTCC CTGCTGCCGT 1920
TTTCACCGA AGTTCATGCC AGTCCAGCGT TTTTGCAGCA GAAAAGCCGC CGACTTCGGT 1980
TTGCGGTCGC GAGTGAAGAT CCCATTCTTG TTACCGCCAA CGCGCAATAT GCCTTGCGAG 2040
GTCGCAAAT CGCCGAAATT CCATACCTGT TCACCGACGA CGGGGCTGAC GCGATCAAAG 2100
ACGCGGTGAT ACATATCCAG CCATGCACAC TGATACTCTT CACTCCACAT GTCGGGTGTAC 2160
ATTGAGTGCA GCCCGGCTAA CGTATCCACG CCGTATTCTGG TGATGATAAT CGGCTGATGC 2220
AGTTTCTCCT GCCAGGCCAG AAGTTCTTT TCCAGTACCT TCTCTGCCGT TTCCAAATCG 2280
CCGCTTGGAA CATACCATCC GTAATAACGG TTCAGGGCACA GCACATCAA GAGATCGCTG 2340
ATGGTATCGG TGTGAGCGTC GCAGAACATT ACATTGACGC AGGTGATCGG ACGCGTCCGG 2400
TCGAGTTAC GCGTTGCTTC CGCCAGTGGC GCGAAATATT CCCGTGCCACC TTGGGGACGG 2460
GTATCCGGTT CGTTGGCAAT ACTCCACATC ACCACGCTTG GGTGGTTTT GTCACGGCCT 2520
34/75

FIG. 8D

ATCAGCTCTT TAATGGCCTG TAAGTGCGCT GGCTGAGTTT CCCCGTTGAC TGCCTCTTCG 2580.
 CTGTACAGTT CTTTCGGCTT GTTGAACCAA TCGAAACCA TGCCCTAAAGA GAGGTTAAAG 2640
 CCGACAGCAG CAGTTTCATC AATCACCACTG ATGCCATGTT CATCTGCCA GTCGAGGCATC 2700
 TCTTCAGCGT AAGGGTAATG CGAGGTTACGG TAGGAGTTGG CCCCAATCCA GTCCATTAAAT 2760
 GCGTGGTCGT GCACCATCAG CACGTTATCG AATCCTTGC CACGCAAGTC CGCATCTTC 2820
 TGACGACCAA AGCCAGTAAA GTAGAACCGGT TTGTGGTTAA TCAGGAACTG TTGCCCTTC 2880
 ACTGCCACTG ACCGGATGCC GACGGGAAGC GGGTAGATAAT CACACTCTGT CTGGCTTTG 2940
 GCTGTGACGC ACAGTTCATC GAGATAACCT TCACCCGGTT GCCAGAGGGTG CGGATTCAACC 3000
 ACTTGCAAAG TCCCCTGCTAGT GCCCTTGTCCA GTTGCAACCA CCTGTTGATC CGCATCACGC 3060
 AGTTCAACGC TGACATCACC ATTGGCCACC ACCTGCCAGT CAACAGACGC GTGGTTACAG 3120
 TCTTGGCGA CATGGCTCAC CACGGTGATA TCGTCCACCC AGGTGTTGG CGTGGTTGTAG 3180
 AGCATTACGC TGGGATGGAT TCCGGCATAG TTAAAGAAAT CATGGAAGTA AGACTGCTTT 3240
 TTCTTGGCGT TTTCGGTGGT AATCACCAATT CCCGGGGGA TAGTCTGCCA GTTCAGTTTCG 3300
 TTGTTCACAC AACGGGTGAT ACCCCTCGAC GGATTAAAGA CTTCAAGGGG TCAACTATGA 3360

35/75

FIG. 8E

AGAAGTGTTC GTCTTCGTCC CAGTAAGCTA TGTCTCCAGA ATGTAGCCAT CCATCCTTGT 3420
CAATCAAGGC GTTGGTCTGCT TTACATAACC GGACATAATC ATAGGTCTC 3480
TGACACATAA TTCGCTCTC TGATTAACGC CGAGCGTTT CCCGGTATCC AGATCCACAA 3540
CCTTCGGCTTC AAAAAATGGA ACAACTTAC CGACCGGGCC CGGTTTATCA TCCCCCTCGG 3600
GTGTAATCAG AATAGCTGTAT GTAGTCTCAG TGAGCCCATA TCCTTGTCTGTT 3660
GATGGAAGGCG TTTTGCAACC GCTTCCCGA CTTCTCGA AAGAGGTGCG CCCCGAGAAG 3720
CAATTTCGGT TAAATTAGAT AAATCGTATT TGTCAATCAG AGTGCCTTTG GCGAAGAATG 3780
AAAATGGGT TGGTACTAGC AACGCCACTT GAATTTGTA ATCCTGAAGG GATCGTAAAA 3840
ACAGCTCTTC TTCAAATCTA TACATTAAGA CGACTCGAAA TCCACATATC AAATATCCGA 3900
GTGTAGTAAA CATTCCAAA CCGTGATGGA ATGGAACAAC ACTTAAAATC GCAGTATCCG 3960
GAATGATTG ATTGCCAAA ATGGATCTC TGGCATGCGA GAATCTAGCG CAGGCAGTTC 4020
TATGCGGAAG GCCCACACCC TTAGGTAACC CAGTAGATCC AGAGGAATTG TTTTGTCACT 4080
ATCAAAGGAC TCTGGTACAA AATCGTATTTC ATTAAAACCG GGAGGTAGAT GAGATGTCAC 4140
GAACGTGTAC ATCGACTGAA ATCCCTGGTA ATCCGTTTA GAATCCATGA TAATAATT 4200

FIG. 8F

CTGGATTATT GGTAATTTTT TTTGCACCGTT CAAAATTTT TGCAAACCCCT TTTTGGAAAC 4260
AACACTACG GTAGGGCTGCG AAATGTTCAT ACTGTTGAGC AATTCACTGTT CATTATAAAT 4320
GTCGTTCGCG GGCGCAACTG CAACTCCGAT AAATAACGCG CCCAACACCCG GCATAAAGAA 4380
TTGAAGAGAG TTTTCACTGCG ATACGACCGAT TCTGTGATT GTATTCAAGCC CATATCGTTT 4440
CATAGCTTCT GCCAACCGAA CGGACATTTC GAAGTATTCC GCGTACGTGA TGTTCACCTC 4500
GATATGTA GCA TCTGTAAAAG GAATTGTTCC AGGAACCAGG GCGTATCTCT TCATAGCCTT 4560
ATGCAGTTGC TCTCCAGCGG TTCCATCCCTC TAGCTTTGCT TCTCAATTTC TTATTTGCAT 4620
AATGAGAAAA AAAGGAAAAT TAATTTAAC ACCAATTTCAG TAGTTGATTG AGCAAATGCG 4680
TTGCCAAAA GGATGCTTA GAGACAGTGT TCTCTGCACA GATAAGGACA AACATTATTC 4740
AGAGGGAGTA CCCAGAGGCTG AGACTCCTAA GCCAGTGAGT GGCACAGCAT TCTAGGGAGA 4800
AATATGCTTG TCATCACCGA AGCCTGATTG CGTAGAGCCA CACCTTGGTA AGGGCCAATC 4860
TGCTCACACA GGATAGAGAG GGCAGGGGCC AGGGCAGAGC ATATAAGGTG AGGTAGGATC 4920
AGTTGCTCCT CACATTGCT TCTGACATAG TTGTGTTGGG AGCTTGGATC GATCCACCAT 4980
GGGCTTCAAT ACCCCTGATTG ACTGGAACAG CTGTAGCCCT GAACAGCAGC GTGCGCTGCT 5040

FIG. 8G

GACGGCTCCG GCGATTTCGG CCTCTGACAG TATTACCCGG ACGGTCAAGC ATATTCTGGAA 5100
TAATGTAAAA ACGCGGGTG ACGATGCCCT GCGTGAATAC AGCGCTAAAT TTGATAAAAC 5160
AGAAGTGACA GCGCTACGGC TCACCCCTGA AGAGATCGCC GCCGCCGGCG CGCGTCTGAG 5220
CGACGAATTA AACAGGGCA TGACCGCTGC CGTCAAAAT ATTGAAACGT TCCATTCCGC 5280
GCAGACGCTA CCGCCTGTAG ATGTGGAAC CCAGCCAGGC GTGCGTTGCC AGCAGGTTAC 5340
GGGTCCCCTGC TCGTCTGTGC GTCTGTATAT TCCCGGGGCC TCGGCTCCGC TCTTCTCAAC 5400
GGTGCTGATG CTGGCGACGC CGGGGGCAT TGCGGGATGC CAGAAGGGTGG TTCTGTGCTC 5460
GCCGCCGCC ATCGCTGATG AAATCCTCTA TGCGGGCAA CTGTGTGCC TGCAAGGAAAT 5520
CTTAAACGTC GGCGGGCGC AGGGGATTGC CGCTCTGGCC TTGGCAGGCC AGTCCCGTACC 5580
GAAAGTGGAT AAAATTNTTG GCCCGGGCAA CGCCTTGTAA ACCGAAGCCA AACGTCAGGGT 5640
CAGCCAGCGT CTCGACGGCG CGGCTATCGA TATGCCAGGC GGGCGTCTG AAGTACTGGT 5700
GATCGCAGAC AGCGGGCAA CACCGGATT CGTCGCTTCT GACTGCTCT CCCAGGCTGA 5760
GCACGGCCG GATTCCAGG TGATCCTGCT GACGGCTGAT GCTGACATTG CCCGCAAGGT 5820
GGCGGAGGGCG GTAGAACGTC AACTGGGGAA ACTGCCGGC GCAGCACCCG CCCGGCAGGC 5880

FIG. 8H

70/41040
PC1/US98/03935

CCTGAGGCC AGTCCGTCTGA TTGTGACCAA AGATTAGCG CAGTGCCTCG CCATCTCTAA 5940
TCAGTATGGG CGGAAACACT TAATCATCCA GACGCCAAT GCGCGCGATT TGGTGGATGC 6000
GATTACCAGC GCAGGCTCGG TATTCTCGG CGACTGGTCG CCGGAATCCG CCGGTGATTA 6060
CGCTTCCGA ACCAACCATG TTTTACCGAC CTATGGCTAT ACTGCTACCT GTTCCAGGCCT 6120
TGGGTTAGG GATTTCCAGA AACGGATGAC CGTTCAAGGAA CTGTCGAAAG CGGGCTTTTC 6180
CGCTCTGGCA TCAACCATTG AAACATTGGC GGCGGCAGAA CGTCTGACCCG CCCATAAAA 6240
TGCCGTGACC CTGCCGGTAA ACCGCCTCAA GGAGCAAGCA TGAGGCACTGA AACACTCTC 6300
AGCGTCGCTG ACTTAGCCCC TGAAAATGTC CGCAACCTGG AGATCCAGAC ATGATAAGAT 6360
ACATTGATGA GTTGGACAA ACCACAACCA GAATGCAGTG AAAAATGCA TTTATTGTG 6420
AAATTGTA TGCTATTGCT TTATTGTA CCATTATAAG CTGCAATAAA CAAGTTAACAA 6480
ACAAACAATTG CATTCAATTG ATGTTTCAGG TTCAGGGGA GGTGTGGAG GTTTTTAA 6540
GCAAGTAAA CCTCTACAAA TGTGGTATGG CTGATTATGA TCTCTAGCTC GACGGCGGC 6600
CTCTAGAGCA GTGTTGGTTT GCAAGAGGAA GCAAAAAGCC TCTCCACCA GGCCTGGAAT 6660
GTTCCACCC AATGTCGAGC AGTGTGGTTT TGCAAGAGGA AGCAAAAGC CTCTCCACCC 6720
39/75

FIG. 8I

AGGCCTGGAA TGTTTCCACC CAATGTCCGAG CAAACCCCGC CCAGCGTCTT GTCATTGGCG 6780
AATCGAACCA CGCAGATGCA GTCGGGGGGG CGCGGTCCCCA GTCCCCACTTC GCATATTAAG 6840
GTGACGGCTG TGGCCTCGAA CACCGAGCGA CCCTGCAGGCC AATATGGGAT CGGCCATTGA 6900
ACAAGATGGA TTGCACCGCAG GTTCTCCGGC CGCTTGGGTG GAGAGGCTAT TCGGCTATGA 6960
CTGGCACAA CAGACAATCG GCTGCTCTGA TGCCGCCGTG TTCCGGCTGT CAGCGCAGGG 7020
GCGCCCGT CTTTTGTCA AGACCGACCT GTCCGGTGCC CTGAATGAAC TGCAGGTAAG 7080
TGCGGCCGTC GATGGCCGAG GCGGCCTCGG CCTCTGCATA AATAAAAAAA ATTAGTCAGC 7140
CATGCATGGG GCGGAGGAATG GGCGGAACTG GGCGGAGTTA GGGCGGGGAT GGGGGAGTT 7200
AGGGGGGA CTATGGTTGC TGACTAATTG AGATGCATGC TTTGCATACT TCTGCCTGCT 7260
GGGGAGGCTG GGGGACTTTCC ACACCTGGTT GCTGACTAAT TGAGATGCAT GCTTTGCATA 7320
CTTCTGCCTG CTGGGGAGCC TGGGGACTTT CCACACCCCTA ACTGACACAC ATTCCACAGA 7380
ATTAATTCCC CTAGTTATTA ATAGTAATCA ATTACGGGGT CATTAGTTCA TAGCCCCAT 7440
ATGGAGTTCC GCGTTACATA ACTTACGGTA AATGGGGGGC CTGGCTGACCC GCGGAAACGAC 7500
CCCCGCCAT TGACGTCAAT AATGACGTAT GTTCCCCATAG TAACGCCAAT AGGGGACTTTTC 7560

40/75

FIG. 8J

CATTGACCGTC AATGGGTGGA CTATTTACGG TAAACTGCC ACTTGGCAGT ACATCAAAGTG 7620
TATCATATGC CAAGTACGCC CCCTATTGAC GTCAATGACG GTAAATGGCC CGCCTGGCAT 7680
TATGCCAGT ACATGACCTT ATGGGACTTT CCTACTTGGC AGTACATCTA GCTATTAGTC 7740
ATCGCTATTA CCATGGTGTGAT GCGGTTTGG CAGTACATCA ATGGGCGTGG ATAGCGGGTT 7800
GACTCACGGG GATTCCAAG TCTCCACCCC ATTGACGTCG ATGGGAGTTT GTTTGGCAC 7860
CAAATCAAC GGGACTTTCC AAAATGTCGT AACAACTCCG CCCCATTGAC GCAAATGGGC 7920
GGTAGGGCTG TAGGGTGGGA GGTCTATA AGCAGAGCTG GGTACGTGAA CCGTCAGATC 7980
GCCTGGAGAC GCCATCACAG ATCTCTCACT ATGGGATTTC AGGTGCAGAT TATCAGCTTC 8040
CTGCTAATCA GTGCTTCAGT CATAATGTCC AGAGGACAAA TTGGTTCTC CCAGTCTCCA 8100
GCAATCCCTGT CTGCATCTCC AGGGGAGAAG GTCACAAATGA CTTGCAGGGC CAGCTCAAGT 8160
GTAAGTTACA TCCACTGGTT CCAGGCAGAAG CCAGGATCCT CCCCCAAACC CTGGATTAT 8220
GCCACATCCA ACCTGGCTTC TGGAGTCCCT GTTCGCTTCA GTGGCAGTGG GTCTGGGACT 8280
TCTTACTCTC TCACAATCAG CAGAGTGGAG GCTGAAGATG CTGCCACTTA TTACTGCCAG 8340
CAGTGGACTA GTAACCCACC CACGTTGGGA GGGGGACCA AGCTGGAAAT CAAACGGTACG 8400

41/75

FIG. 8K

GTGGCTGCAC CATCTGTCTT CACCTCCCCG CCATCTGATG AGCAGTTGAA ATCTGGAACT 8460
GCCCTCTGTTG TGTGCTGT GAATAACTTC TATCCCAAGAG AGGCCAAAGT ACAGTGGAAAG 8520
GTGGATAACG CCCTCCAATC GGGTAACTCC CAGGAGAGTG TCACAGAGCA GGACAGCAAG 8580
GACAGCACCT ACAGCCTCAG CAGCACCCCTG ACGCTGAGCA AAGCAGACTA CGAGAAACAC 8640
AAAGTCTACG CCTGCGAAGT CACCCATCAG GGCCTGAGCT CGCCCGTCAC AAAGAGCTTC 8700
AACAGGGAG AGTGTGAAT TCAGATCCGT TAACGGTTAC CAACTACCTA GACTGGATTTC 8760
GTGACAACAT GCGGGCCGTGA TATCTACGTA TGATCAGGCCT CGACTGTGCC TTCTAGTTGC 8820
CAGCCATCTG TTGTTGCC CTCCCCGGTGA CCTTCCCTTGA CCCTGGAAAGG TGCCACTCCC 8880
ACTGTCCCTT CCTAATAAAA TGAGGAAATT GCATCGCATT GTCTGAGTAG GTGTCAATTCT 8940
ATTCTGGGG GTGGGGTGGG GCAGGACAGC AAGGGGGAGG ATTGGGAAGA CAATAGCAGG 9000
CATGCTGGGG ATGCCGTGGG CTCTATGGAA CCAGCTGGG CTCGACAGCT ATGCCAAGTA 9060
CGCCCCCTAT TGACGTCAAAT GACGGTAAAT GCCCTGGCCTG GCATTATGCC CAGTACATGA 9120
CCTTATGGGA CTTTCTACT TGGCAGTACA TCTACGTATT AGTCATCCGCT ATTACCATGG 9180
TGATGCGGTT TTGGCAGTAC ATCAAATGGGC GTGGATAGCC GTTTGACTCA CGGGGATTTC 9240

FIG. 8L

43/75

CAAGTCTCCA CCCCATTGAC GTCAATGGGA GTTTGTTCG GCACCCAAAT CAACGGGACT 9300
 TTCCAAATG TCGTAACAAC TCCGCCCAT TGACGCAAAT GGGCGGTAGG CGTGTACGGT 9360
 GGGAGGTCTA TATAAGCAGA GCTGGGTACG TCCTCACATT CAGTGATCAG CACTGAACAC 9420
 AGACCCGTCG ACATGGGTTG GAGCCTCATC TTGCTCTTCC TTGTCGCTGT TGCTACGGGT 9480
 GTCTGTCCC AGGTACAAC GCAGGCAGCCT GGGGCTGAGC TGGTGAAGGCC TGGGGCCTCA 9540
 GTGAAGATGT CCTGCAAGGC TTCTGGCTAC ACATTTACCA GTTACAATAT GCACTGGTA 9600
 AACAGACAC CTGGTGGGG CCTGGAATGG ATTGGAGCTA TTTATCCGG AAATGGTGAT 9660
 ACTTCCTACA ATCAGAAGTT CAAAGGCAAG GCCACATTGA CTGCAGACAA ATCCTCCAGC 9720
 ACAGCCTACA TGCAGCTCAG CAGCCTGACA TCTGAGGGACT CTGCGGTCTA TTACTGTGCA 9780
 AGATCGACTT ACTACGGGG TGACTGGTAC TTCAATGTCT GGGGGCAGG GACCACGGTC 9840
 ACCGTCTG CAGCTAGCAC CAAGGGCCA TCGGTCTTCC CCCTGGCACCC CTCCTCCAAG 9900
 AGCACCTCTG GGGGCACAGC GGCCCTGGGC TGCGTGGTCA AGGACTACTT CCCCCGAACCG 9960
 GTGACGGGTGT CGTGGAACTC AGGGGCCCTG ACCAGGGGGCG TGCACACCTT CCGGGCTGTC 10020
 CTACAGTCCT CAGGACTCTA CTCCCCTCAGC AGCGTGGTGA CCGTGCCTC CAGCAGCTTG 10080

FIG. 8M

GGCACCCAGA CCTACATCTG CAACGTGAAT CACAAGCCAA GGTGGACAAG 10140
AAAGCAGAGC CCAAATCTTG TGACAAACTCACACATGCC CACCGTGCC AGCACCTGAA 10200
CTCCTGGGG GACCGTCAGT CTTCCCTTTC CCCCAAAAC CCAAGGACAC CCTCATGATC 10260
TCCGGACCC CTGAGGTAC ACAGCACGTA CGGTGTGGTC AGCGTCCCTCA CCAGGACTGG 10320
AAGTTCAACT GGTACCGTGGAA CGGGCGTGGAG GTGCATAATG CCAAGACAAA GCCGGGGAG 10380
GAGCAGTACA AGGAGTACAA GTGCAAGGTC TCCAACAAAG CCCTCCCAGC CCCCATCGAG 10440
CTGAATGGCA AGGAGTACAA GTGCAAGGTC TCCAACAAAG CCCTCCCAGC CCCCATCGAG 10500
AAACCATCT CCAAAGCCAA AGGGCAGCCC CGAGAACACC AGGTGTACAC CCTGCCCCA 10560
TCCGGGATG AGCTGACCA GAACCAGGTC AGCCTGACCT GCCTGGTCAA AGGCTTCTAT 10620
CCCAGGGACA TGGCGTGGA GTGGGAGGAGC AATGGGGAGC CGGAGAACAA CTACAAGACC 10680
ACGGCTCCCG TGCTGGACTC CGACGGCTCC TTCTTCCTCT ACAGCAAGCT CACCGTGGAC 10740
AAGAGCAGGT GGCAGCAGGG GAAACGTCTTC TCATGCTCCG TGATGCATGA GGCTCTGCAC 10800
AACCACTACA CGCAGAAGAG CCTCTCCCTG TCTCCCCCTG AATGAGGATC CGTTAACGGT 10860
TACCAACTAC CTAGACTGGAA TTCCGTGACAA CATGCGGGCCG TGATATCTAC GTATGATCAG 10920

44/75

FIG. 8N

CCTCGACTGT GCCTTCTAGT TGCCAGCCAT CTGTTGTTG CCCACTGTCC TTTCCTAATA AAATGAGGAA ATTGCATCGC 10980
TGACCCCTGGA AGGTGCCACT CCCACTGTCC TTTCCTAATA AAATGAGGAA ATTGCATCGC 11040
ATTGCTGAG TAGGTGTCAT TCTATTCTGG GGGGTGGGT GGGCAGGGAC AGCAAGGGGG 11100
AGGATTGGGA AGACAATAGC AGGCATGCTG GGGATGCGGT GGGCTCTATG GAACCAGGCTG 11160
GGGCTCGACA GCAACGCTAG GTCGAGGCCG CTACTAACTC TCTCCTCCCT CCTTTTCCCT 11220
GCAGGACGAG GCAGCGGGC TATCGTGGCT GGCCACGACG GGCGTTCCCT GCGCAGGCTGT 11280
GCTCGACGTT GTCACTGAAG CGGGAAGGGAA CTGGCTGCTA TTGGCGGAAG TGCCGGGCA 11340
GGATCTCCTG TCATCTCACC TTGCTCCTGC CGAGAAAAGTA TCCATCATGG CTGATGCAAT 11400
CGGGCGGCTG CATAAGCTTG ATCCGGCTAC CTGCCATTG GACCAACAG CGAACATATCG 11460
CATCGAGCGA GCACGTTACTC GGATGGAAAGC CGGTCTTGTG GATCAGGGATG ATCTGGACGA 11520
AGAGCATCAG GGGCTCGCGC CAGCCGAACG GTTCGCCAGG TAAGTGAGCT CCAATTCAAG 11580
CTTCCTAGGG CGGCCAGGCTA GTAGCTTGC TTCTCAATT CTTATTTGCA TAATGAGAAA 11640
AAAAGGAAA TTAATTAA CACCAATTCA GTAGTTGATT GAGCAAATGCA GTTGCCAAA 11700
AGGATGCTTT AGAGACAGTG TTCTCTGCAC AGATAAGGAC AACACATTATT CAGAGGGAGT 11760

45/75

FIG. 8P

ACCCAGAGCT GAGACTCCTA AGGCCAGTGAG, TGGCACAGCA TTCTAGGGAG AAATATGCTT 11820
GTCATCACCG AAGCCTGATT CCGTAGAGCC ACACCTTGGT AAGGGCCAAT CTGCTCACAC 11880
AGGATAGAGA GGGCAGGAGC CAGGGCAGAG CATATAAGGT GAGGTAGGAT CAGTTGCTCC 11940
TCACATTGC TTCTGACATA GTTGTGTTGG GAGCTTGGAT AGCTTGGACA GCTCAGGGCT 12000
GGGATTTCGC GCCAAACTTG ACGGCAATCC TAGCGGTGAAG GCTGGTAGGA TTTTATCCCC 12060
GCTGCCATCA TGGTTCGACC ATTGAAC TGTCGATCC ATCGTCGCC TGTCCTCGGCTTAAA TATGGGGATT 12120
GGCAAGAACG GAGACCTACC CTGGCCTCCG CTCAGGAACG AGTTCAAGTA CTTCCAAAGA 12180
ATGACCAAA CCTCTTCAGT GGAAAGGTAAA CAGAATCTGG TGATTATGG TAGGAAAACC 12240
TGGTTCTCCA TTCCCTGAGAA GAATCGACCT TTAAAGGACA GAATTAAATAT AGTTCTCAGT 12300
AGAGAACTCA AAGAACCC ACCGGGAGCT CATTTCCTTG CCAAAAGTTT GGATGATGCC 12360
TTAAGACTTA TTGAACAAACC GGAAATTGGCA AGTAAAGTAG ACATGGTTG GATACTGGGA 12420
GGCAGTTCTG TTTACCGGA AGCCATGAAT CAACCAAGGCC ACCCTTAGACT CTTTGTGACA 12480
AGGATCATGC AGGAATTGA AAGTGACACCG TTTCCTCCAG AAATTGATTT GGGAAATAT 12540
AAACTTCTCC CAGAATAACCC AGGCCTCCTC TCTGAGGTCC AGGAGGAAA AGGCATCAAG 12600

46/75

FIG. 8Q

TATAAGTTG AAGTCTACGA GAAGAAAGAC TAACAGGAAG ATGCTTTCAA GTTCTCTGCT 12660
CCCTCCTAA AGCTATGCAT TTTATAAGA CCATGGACT TTTGCTGGCT TTAGATCAGC 12720
CTCGACTGTG CCTTCAGTT GCCAGCCATC TGTGTTGC CCCTCCCCG TGCCTTCCTT 12780
GACCCTGGAA GGTGCCACTC CCACTGTCCCT TTCCCTAATAA AATGAGGAAA TTGCATCGCA 12840
TTGTCTGAGT AGGTGTCATT CTATTCTGGG GGGTGGGGTG GGGCAGGACA GCAAGGGGA 12900
GGATTGGAA GACAATAGCA GGCATGCTGG GGATGCGGTG GGCTCTATGG AACCAGCTGG 12960
GGCTCGAAGC GGCGCCCAT TTGCTGGTG GTCAGATGCG GGATGGCGTG GGACGGGGCG 13020
GGGAGGGTCA CACTGAGGTT TTCCGCCAGA CGCCCACTGCT GCCAGGGCCT GATGTGCCCG 13080
GCTTCTGACC ATGCGGTGCG GTTCGGTTGC ACTACGGCTA CTGTGAGCCA GAGTTGCCCG 13140
GGGCTCTGG GCTGGGTAG TTCAATCAACT GTTACCTTG TGACCCGACA 13200
TCCAGAGGCA CTTCACCGCT TGCCAGGGC TTACCATCCA GCGCCACCAT CCAGTGCAGG 13260
AGCTCGTTAT CGCTATGACG GAACAGGTAT TCGCTGGTCA CTTCGATGGT TTGCCGGAT 13320
AACGGAACT GGAAAAACTG CTGCTGGTGT TTTGCTTCGG TCAGGGCTGG ATGGGGGTG 13380
CGGTGGCAA AGACCAAGACG GTTCATACAG AACTGGCGAT CGTTCGGGCGT ATCGCCAAA 13440

47/75

FIG. 8R

TCACCGCCGT AAGCCGACCA CGGGTTGCCG TTTCATCAT ATTAAATCAG CGACTGATCC 13500
 ACCCAGTCCC AGACGAAGCC GCCCTGTAAA CGGGATACT GACGAAACCGC CTGCCAGTAT 13560
 TTAGC GAAAC CGCCAAGACT GTTACCCATC GCTGGGGCGT ATT CGCAAAG GATCAGCGGG 13620
 CGCGTCTC CGGGTAGCGA AAGCCATT TTGATGGACC ATTCGCAAAG AGCCGGGAAG 13680
 GGCTGGTCTT CATCCACCG CGCGTACATC GGGCAAATAA TATCGGTGGC CGTGGTGTGCG 13740
 GCTCCGCC CTTCATACTG CACCGGGCGG GAAGGATCGA CAGATTGAT CCAGCGATAAC 13800
 AGCGCGT GATTAGGCC GTGGCCTGAT TCATTCCCCA GCGACCAAGAT GATCACACTC 13860
 GGGT GATTAC GATCGCGCTG CACCATT CGTACCGCGTT CGCTCATCGC CGGTAGGCCAG 13920
 CGCGGATCAT CGGT CAGACG ATT CATTGGC ACCATGCCGT GGGTTCAAT ATTGGCTTCA 13980
 TCCACCACAT ACAGGGCGTA GCGGTCCAC AGCGTGTACC ACAGGGGATG GTTGGATAA 14040
 TGCCAACAGC GCACGGCGTT AAAGTTGTTC TGCTTCATCA GCAGGGATATC CTGCACCATC 14100
 GTCTGCTCAT CCATGACCTG ACCATGCAGA GGATGATGCT CGTGACGGTT AACGCCCTCGA 14160
 ATCAGCAACG GCTTGGCGTT CAGCAGGAGC AGACCAATT CAATCCGCAC CTCGGGAAA 14220
 CCGACATCGC AGGCTTCTGC TTC AATCAGC GTGCCGGCGG CGGTGTGCAG TTCAACACC 14280

48/75

FIG. 8S

GCACGATAGA GATTGGGGAT TTCCGGCGCTC CACAGTTTCG GGTTTTCGAC GTTCAGACGC 14340
 AGTGTGACGC GATCGGCATA ACCACCACGC TCATCGATAA TTTCACCGGCC GAAAGGGCGCG 14400
 GTGCCGCTGG CGACCTGCGT TTCACCCCTGC CATAAAGAAA CTGTTACCCG TAGGTAGTCA 14460
 CGCAACTCGC CGCACCATCTG AACTTCAGGCC TCCAGTACAG CGCGGGCTGAA ATCATCATT 14520
 AAGCGAGTGG CAACATGGAA ATCGCTGATT TGTGTAGTCG GTTATGCAG CAACGAGACG 14580
 TCACGGAAA TGCCGCTCAT CGGCCACATA TCCTGATCTT CCAGATAACT GCCGTCACTC 14640
 CAACGCAGCA CCATCACCGC GAGGGGGTTT TCTCCGGCGC GTAAAAATGC GCTCAGGTCA 14700
 AATTCAAGACG GCAAACGACT GTCCCTGGCCG TAACCGACCC ACGCCCCGTT GCACCCACAGA 14760
 TGAAACGCC AGTTAACGCC ATCAAAATA ATTTCGGTCT GGCCCTTCCCTG TAGCCAGCTT 14820
 TCATCAACAT TAAATGTGAG CGAGTAACAA CCCGTGGGAT TCTCCGTGGG AACAAACGGC 14880
 GGATTGACCG TAATGGATA GGTTACGTTG GTGTAGATGG GCGCATCGTA ACCGTGCATC 14940
 TGCCAGTTG AGGGGACGAC GACAGTATCG GCCTCAGGAA GATCGCACTC CAGCCAGCTT 15000
 TCCGGCACCG CTTCTGGTGC CGGAAACCAAG GCAAAGGGCC ATTCCGCATT CAGGCTGGCC 15060
 AACTGTTGGG AAGGGCGATC GGTGGGGGCC TCTTCGCTAT TACGCCAGCT GGCGAAAGGG 15120

49/75

FIG. 8T

GGATGTGCTG CAAGGCCATT AAGTTGGTA ACGGCCAGGGT TTTCCCAGTC ACGACGGTTGT 15180
AAACGACTT ATCCGTCGA GGGGCTGCCT CGAAGCAGAC GACCTTCCGT TGTGCAGGCCA 15240
GCGCGCCTG CGCCGGTGCCT CACAATCGTG CGCGAACAAA CTAAACCAGA ACAAATTATA 15300
CCGGCGCAC CGCCGCCACC ACCCTCTCCC GTGCCCTAACCA TTCCAGCGCC TCCACCCACCA 15360
CCACCCAT CGATGTCTGA ATTGCCGGCC GCTCCACCAA TGCCGACGGA ACCTCAACCC 15420
GCTGCACCT TAGACGACAG ACAACAATTG TTGGAAGCTA TTAGAAACGA AAAAAATCGC 15480
ACTCGTCTCA GACCGGTCAA ACCAAAAACG GGGCCGGAAA CCAGTACAAT AGTTGAGGGTG 15540
CCGACTGTGT TGCCCTAAAGA GACATTTGAG CCTAAACCGC CGTCTGCATC ACCGCCACCA 15600
CCTCCGCC CGCCTCCGCC GCCAGCCCCG CCTGCCCTC CACCGATGGT AGATTTATCA 15660
TCAGCTCCAC CACCCGCC ATTAGTAGAT TTGCCGTCTG AAATGTTACC ACCGCCCTGCA 15720
CCATCGCTT CTAACCGTGT GTCTGAATT AAATCGGGCA CAGTTAGATT GAAACCCGCC 15780
CAAAACGCC CGCAATCAGA AATAATTCCA AAAAGCTCAA CTACAAATTG GATCGGGGAC 15840
GTGTTAGCCG ACACAAATTAA TAGGGCTATGG CAAAATCGTC TTGGGAAGCA 15900
ACTTCTAACG ACGAGGGTTG GGACGACGAC GATAATCGGC CTAATAAAGC TAACACGCC 15960

50/75

FIG. 8U

GATGTTAAAT ATGTCCAAGC TACTAGTGGT ACCGGCTTGGC AGAACATATC CATCGCGTCC 16020
 GCCATCTCCA GCAGCCGCAC CGGGCGCATC TCGGGCAGCG TTGGGTCCCTG GCCACGGGTG 16080
 CGCATGATCG TGCTCCTGTC GTTGAGGACC CGGCTAGGCT GGGGGGGGTGC CCTTACTGGT 16140
 TAGCAGAATG AATCACCGAT ACGCGAGCGA ACGTGAAGCG ACTGCTGCTG CAAAACGTCT 16200
 GCGACCTGAG CAACAACATG AATGGTCTTC GGTTCCCGTG TTTCGTAAG TCTGGAAACG 16260
 CGGAAGTCAG CGCCCTGCAC CATTATGTTTC CGGATCTGCA TCGCAGGATG CTGCTGGCTA 16320
 CCCTGTGAA CACCTACATC TGTATTAACG AAGCGCTGGC ATTGACCCCTG AGTGATTTT 16380
 CTCTGGTCCC GCCGCATCCA TACCGCCAGT TGTTTACCCCT CACAAACGTTTC CAGTAACCGG 16440
 GCATGTTCAT CATCAGTAAC CCGTATCGTG AGCATCCTCT CTCGTTTCAT CGGTATCATT 16500
 ACCCCCATGA ACAGAAATCC CCCTTACACCG GAGGCATCAG TGACCAAACA GGAAAAAAC 16560
 GCCCTTAACA TGGCCCGCTT TATCAGAAGC CAGACATTAA CGCTTCTGGA GAAACTCAAC 16620
 GAGCTGGACG CGGATGAACA GGCAGACATC TGTGAATCGC TTCACGACCA CGCTGATGAG 16680
 CTTACCGCA GCTGCCCTCGC GCGTTCCGGT GATGACGGGT AAAACCTCTG ACACATGGCAG 16740
 CTCCGGAGA CGGTACACAGC TTGTCCTGTAAG CCGGATGCCG GGAGCAGACA AGCCCGTCAG 16800

51/75

FIG. 8V

GGCGCGTCAG CGGGTGTGG CGGGTGTGG CGGGCAGCCA TGACCCAGTC ACGTAGCGAT 16860
AGCGGAGTGT ATACTGGCTT AACTATGGGG CATCAGAGCA GATTGTACTG AGAGTGCACC 16920
ATATGCGGTG TGAAATAACCG CACAGATGCG TAAGGAGAAA ATACCGCATC AGGCCGCTTT 16980
CGGCTTCCTC GCTCACTGAC TCGGTGCGCT CGGTGGCTCG GCTGGGGCGA GCGGTATCAG 17040
CTCACTCAA GGCGGTAATA CGGTTATCCA CAGAACATCAGG GGATAACGCA GGAAAGAACAA 17100
TGTGAGCAA AGGCCAGCAA AAGGCCAGGA ACCGTAaaaa GGCGCGGTGG CTGGCGTTT 17160
TCCATAGGCT CGGGCCCCCT GACGAGGCATC ACAAAATCG ACGCTCAAGT CAGAGGTGGC 17220
GAAACCCGAC AGGACTATAA AGATACCAGG CGTTTCCCCC TGGAAAGCTCC CTCGTGCGCT 17280
CTCCTGTTCC GACCCCTGCCG CTTACCGGAT ACCTGTCCGC CTTTCTCCCT TCAGGGAAAGCG 17340
TGGCGCTTTC TCATAGCTCA CGCTGTAGGT ATCTCAGTTTC GGTGTAGGTG GTTCCGCTCCA 17400
AGCTGGGCTG TGTGCACGAA CCCCGGTTTC AGCCCCGACCC CTGCGCCCTTA TCCGGTAACT 17460
ATCGTCTTGA GTCCAACCCG GTAAGACACG ACTTATCGCC ACTGGCAGCA GCCCACTGGTA 17520
ACAGGATTAG CAGAGGGAGG TATGTAGGCG GTGCTACAGA GTTCTTGAAAG TGGTGGCCTA 17580
ACTACGGCTA CACTAGAAGG ACAGTATTG GTATCTGCGC TCTGCTGAAG CCAGTTACCT 17640

52/75

FIG. 8W

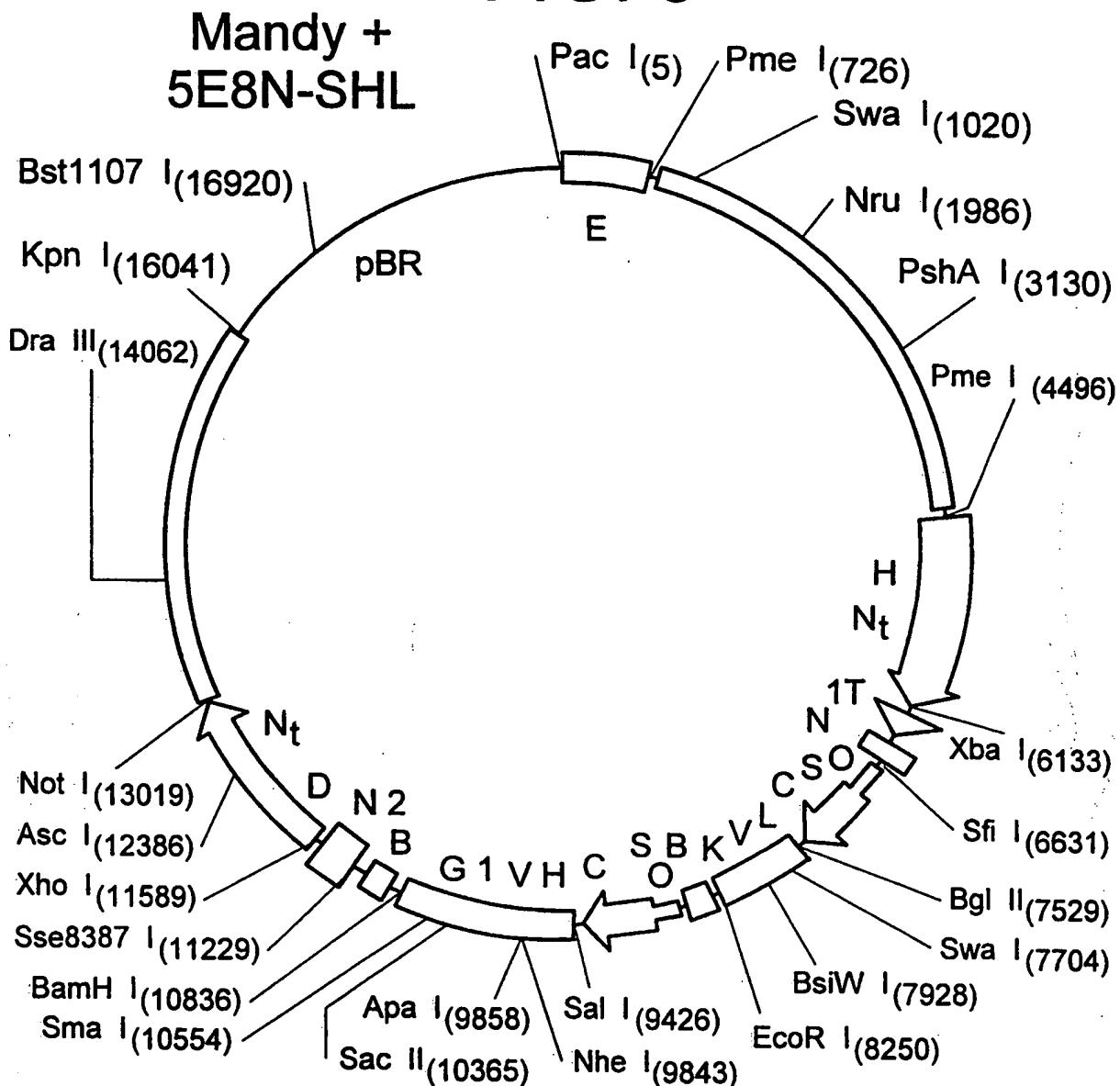
TCGGAAAAAG AGTTGGTAGGC TCTTGATCCG GCAAACAAAC CACCGCTGGT AGCGGGTGGTT 17700
TTTTGTTTG CAAGCAGCAG ATTACGCGCA GAAAAAAAGG ATCTCAAGAA GATCCTTTGA 17760
TCTTTCTAC GGGGTCTGAC GCTCAGTGGA ACGAAAACCT ACGTTAAGGG ATTTTGGTCA 17820
TGAGATTATC AAAAGGATC TTACACCTAGA TCCTTTAAA TTAAAATGA AGTTTTAAAT 17880
CAATCTAAAG TATATATGAG TAAACTTGGT CTGACAGTTA CCAATGCTTA ATCAGTGAGG 17940
CACCTATCTC AGCGATCTGT CTATTTCGTT CATCCATAGT TGCCCTGACTC CCCGTCGTGT 18000
AGATAACTAC GATAACGGGAG GGCTTACCAT CTGGCCCCAG TGCTGCAATG ATACCGGAG 18060
ACCCACGCTC ACCGGCTCCA GATTATCAG CAATAAACCA GCCAGCCGGA AGGGCCGAGC 18120
GCAGAAGITGG TCCTGCAACT TTATCCGCCT CCATCCAGTC TATTAATTGT TGCCGGGAAG 18180
CTAGACTAAG TAGTTGCCA GTTAATAGTT TGGCAACGT TGTGCCATT GCTGCAGGCA 18240
TCGTGGTGTAC ACGCTCGTCA TTTGGTATGG CTTCATTCAAG CTCCGGTTCC CAACGATCAA 18300
GGCGAGTTAC ATGATCCCCC ATGTTGTGCA AAAAGGGT TAGCTCCTTC GGCTCCTCGA 18360
TCGTTGTCAG AAGTAAGTTG GCCGGCAGTGT TATCACTCAT GGTTATGGCA GCACCTGCATA 18420
ATTCTCTTAC TGTCAATGCCA TCCGTAAGAT GCTTTTCTGT GACTGGTGAG TACTCAACCA 18480
53/75

FIG. 8X

AGTCATTCTG AGAATAGTGT ATGCCGGAC CGAGTTGCTC TTGCCCGGCG TCAACACGGG 18540
ATAATCCGC GCCCACATAGC AGAACTTAA AAGTGCTCAT CATTGGAAA CGTTCTCGG 18600
GGCGAAACT CTCAAGGATC TTACCGCTGT TGAGATCCAG TTGGATGTA CCCACTCGTG 18660
CACCCAACGT ATCTTCAGCA TCTTTACCT TCACCGCGT TTCTGGGTGA GCAAAACAG 18720
GAAGGCAAAA TGCCGCAAAA AAGGGAAATAA GGGCGACACG GAAATGTTGA ATACTCATAC 18780
TCTCCTTT TCAATATTAT TGAAGCATT ATCAGGGTTA TTGTCTCATG AGCGGATACA 18840
TATTGAATG TATTTAGAAA AATAAACAAA TAGGGGTTCC GGGCACATT CCCGAAAAG 18900
TGCCACCTGA CGTCTAAGAA ACCATTATTA TCATGACATT AACCTATAA AATAGGCGTA 18960
TCACGAGGCC CTTCTGTCTT CAAGAA 18986

54/75

55/75
FIG. 9



Nt D = Inactive Dihydrofolate reductase

E = CMV and SV40 enhancers

Nt H = Inactive *Salmonella* Histidinol Dehydrogenase

T = Herpes Simplex thymidine kinase promoter and polyoma enhancer

C = Cytomegalovirus promoter/enhancer

N1 = Neomycin phosphotransferase exon 1

K = Human kappa constant

VL = Variable light chain anti-CD23 primate 5E8 and leader

VH = Variable heavy chain anti-CD-23 primate 5E8N- and leader

B = Bovine growth hormone polyadenylation

M2 = Neomycin phosphotransferase exon 2

G1 = Human Gamma 1 constant

Mandy cut XbaI Xho I and ligated to Xba I Xho I fragment

from XKG1+CD23 5E8N-SHL

Map by Mitchell Reff Constructed by Karen McLachlan 06/26/97 19,035 bp

Noncutters = AflII, AvrII, HindIII, I-Ppol, I-SceI, PmlI, RsrII, SgfI, SrfI

56/75

10	TTAATTAAAGG	20	GGCGGAGAAT	30	GGCGGGAAC	40	GGGGGGAGTT	50	AGGGGGGGGA	60	TGGGGGGAGT	70	TAGGGGGGGG
80	ACTATGGTTG	90	CTGACTAATT	100	GAGATGCATG	110	CTTTCGCATAC	120	TTCCTGCCTGC	130	GGGGAGGCCT	140	GGGGACTTTTC
150	CACACCTGGT	160	TGCTGACTAA	170	TTGAGATGCA	180	TGCTTGCAT	190	ACTTCTGCCT	200	GCTGGGGAGC	210	CTGGGGACTT
220	TCCACACCT	230	AACTGACACA	240	CATTCCACAG	250	AATTAATTCC	260	CCTAGTTATT	270	AATAGTAATC	280	AATTACCGGG
290	TCATTAGTTC	300	ATAGCCCATA	310	TATGGAGTTC	320	CGCGTTACAT	330	AACCTTACGGT	340	AAATGGCCCC	350	CCTGGCTGAC
360	CGCCCAACGAA	370	CCCCGGCCA	380	TTGACGTCAA	390	TAATGACGTA	400	TGTTCCCATA	410	TAGGGACTTT	420	TAGGGACTTT
430	CCATTGACGT	440	CAATGGGTGG	450	AGTATTACG	460	GTAAACTGCC	470	CACCTGGCAG	480	GTATCATATG	490	GTATCATATG
500	CCAAGTACGC	510	CCCTTATTGA	520	CGTCAATGAC	530	GGTAAATGGC	540	TACATCAAGT	550	TACATGACCT	560	TACATGACCT
570	TATGGACTT	580	TCCTACTGG	590	CAGTACATCT	600	ACGTATTAGT	610	TTATGCCAG	620	TGCGGGTTTG	630	TGCGGGTTTG
640	GCAGTACATC	650	AATGGCGGTG	660	TGACTCAGG	670	CATGCTATT	680	ACCATGGTGA	690	CATTGACGTC	700	CATTGACGTC
710	AATGGGAGTT	720	TGTTTGAAG	730	GATAGGGTT	740	TGACTCACGG	750	TTTAACGTGT	760	TTACGTGAG	770	TTACGTGAG
780	TCAATTGTAC	860	ACTAACGACA	870	CTGTTAAC	880	GGCAGCTTA	890	GTCTCCACCC	900	CGTCGAACCT	910	CGTCGAACCT
850	TTATTACAAA	930	ACAAAACACA	940	AACGAATATC	950	AAATACAAAA	960	TTTGAAACGA	970	GCAAGATTG	980	GCAAGATTG
920	TGGCGTTGAG		CGAAAATCCA		TTAGATAAGTC		CAGCCATCGG		TTCCGGAAAAA		CAACCCCTGT		TTGAAACTAA

FIG. 10A

57/75

990	TCGAAACCTA	1000	TTTACAAAT	1010	CTATTGAGGA	1020	TTAACATATT	1030	AAATTCAAGAT	1040	TGAAAATCAGC	1050	
1060	TTGATTTCG	1070	CTCTAACATA	1080	CCACCCCTAAA	1090	GATTATAAAT	1100	TTAATGAATT	1110	ATCAGCAACT	1120	
1130	ATATATTGAT	1140	AGACATTTCC	1150	AGTTTGTGAT	1160	ATTAGTTTGT	1170	GGTCTCATT	1180	TTATTTTAA	1190	
1200	CAACAAACAA	1210	CTGCTCGCAG	1220	ACAATAGTAT	1230	AGAAAAGGGAA	1240	GGTGAACTGT	1250	CGGTTCGTAC	1260	
1270	AACATTTGG	1280	AAAGTTATGT	1290	TAATCCGGTG	1300	CTGCTAAAAA	1310	TTTGTGTTAA	1320	GAAGCTGCCG	1330	
1340	ACTATGCCGG	1350	CAACATATTG	1360	TACAAAACCG	1370	ACGATCCCAA	1380	TGAACTAGAA	1390	TAATAATTAA	1400	
1410	AGCAACACAC	1420	TCCGAAGAAC	1430	TACCAGAAAAA	1440	TAGCACTGTT	1450	GTAAATTACA	1460	GCGCAGCGGT	1470	
1480	ACTATACACC	1490	CCATTAAAAAA	1500	AGACATATAT	1510	ATTATGACA	1520	TACTCTATAC	1530	GATAGATAACA	1540	
1550	TATATGGATA	1560	CGATAAAC	1570	TATGTTAATT	1580	TTTATGAGGA	1590	AAAAATGAA	1600	AATACGAAGA	1610	
1620	AGAAGACGAC	1630	AAGGGCTCTA	1640	GTTTATGTGA	1650	AAATAAAATT	1660	AAATTAACGT	1670	TGAATCATT	1680	
1690	GAAAATGATT	1700	TTAAATATTA	1710	CCTCAGCGAT	1720	TATAACTACG	1730	ATATTGTCGC	1740	1750		
1760	TTCTTGTGC	1770	GTTGGTTTG	1780	TATCGTTAAT	1790	AAAAAACAAA	1800	TATAGATAAT	1810	ACTACAAATG		
1830	TAATTACAAA	1840	ACCGCTTGAG	1850	TTGCCAGCAA	1860	TTGACATT	1870	ATAATTGTTT	1880	TATTATTCAA	1890	
1900	ATTCAATTGTT		TAGGATTGAG	1910	1920	TCACCGAAGT	1930	ACGGACAGAG	1940	CTTGTCCGAGG	1950	AGAGTTGTTG	1960
			TGCCTCCCTG		CTGGGGTTT		TCATGCCAGT		CCAGCGTTT		TGCAGCAGAA		

FIG. 10B

58/75

1970	AAGCCGCCGA	C T TCGGTTG	1980	1990	CGGTCGGCGAG	TGAAGATCCC	2000	TTTCTTGTAA	CGGCCAACGCC	2020	GCAATATGCC
2040	TTGGAGGTC	2050	2060	2070	CGAACATTCCA	TACCTGTTCA	2080	CGGACGACGG	CGCTGACGCG	2090	ATCAAAGACG
2110	CGGTGATACA	2120	2130	2140	TGCACACTGA	TACTCTTCAC	2150	TCCACATGTC	GGTGTACATT	2160	GAGTGCAGCC
2180	CGGCTAACGT	2190	2200	2210	TATTGGTGA	TGATAATCGG	2220	CTGATGCAGT	TTCTCCTGCC	2230	AGGCCAGAAG
2250	TTCTTTTCC	2260	2270	2280	CTGCCGTTTC	CAAATCGCCG	2290	CTTGGGACAT	ACCATCCGTA	2300	ATAACGGTTC
2320	AGTACCTTCT	2330	2340	2350	GTATGGTGT	GAGCGTCGCA	2360	GAACATTACA	TTGACGCCAGG	2380	2310
2390	AGGCACAGCA	2400	2410	2420	ATCGCTGATG	GTATGGTGT	2430	CAGTGGCGCG	AAATATTCCC	2440	GTGACCTTG
2460	TGATCGGACG	2470	2480	2490	AGTTTACGCG	TTGCTCCGC	2500	ACGCTTGGGT	GGTTTTGTC	2510	2520
2530	CGGACGGTA	2540	2550	2560	TGGCAATTACT	CCACATCACC	2570	CTCTGACTGC	CTCTTGGCTG	2580	ACGGCTATC
2600	AGCTCTTAA	2610	2620	2630	TGGCCTGTA	TGAGTTCCC	2640	TACAGTTCTT	TTTCAAT	2650	2590
2670	TCGGCTTGT	2680	2690	2700	GTGCCGCTTCG	CTAAAGAGAG	2710	ACAGCAGCAG	ACAGCAGCAG	2720	2730
2740	CACCAAGATG	2750	2760	2770	AAACCAATGC	GAGGATCTCT	2780	GGTAATGCGA	GGTACGGTAG	2790	2800
2810	GAGTTGGCCC	2820	2830	2840	CTGCCCCAGTC	TGGTCGTGCA	2850	GTATTATGAAAT	CCTTTGCCAC	2860	2870
2880	GCAAGTCCGC	2890	2900	2910	CATTAATGCG	CCATCAGCAC	2920	TGGTTAAATCA	GGAAACTGTTG	2930	2940
	GCCCTTCACT				GGATGCGAC	GCGAAGGCCGG		TAGATATCAC	ACTCTGTCTG		

FIG. 10C

59/75

2950	GTGACGCACA	2960	GTTCATAGAG	2970	ATAACCTTCA	2980	CCCGGTTGCC	2990	AGAGGGTGGG	3000	TGCAAAGTCC
3020	CGCTAGTGGC	3030	GCAACCACCT	3040	GTTGATCCGG	3050	ATCACGGCAGT	3060	TCAACGGCTGA	3070	CATCACCATT
3090	GGCCACCACC	3100	TGCCAGTCAA	3110	CAGACGCGTG	3120	GTACAGTCT	3130	TGGCGACAT	3140	GGTGATATCG
3160	TCCACCCAGG	3170	TGTTGGCGT	3180	GGTAGAGGC	3190	ATTACGCTGC	3200	GATGGATTCC	3210	AAGAAATCAT
3230	GGAAAGTAAGA	3240	CTGCTTTTC	3250	TTGCCGTTTT	3260	CGTGGTAAT	3270	GGCATAGTTA	3280	TCTGCCAGTT
3300	CAGTTCGTTG	3310	TTCACACAAA	3320	GGTGATACC	3330	CCTCGACGGA	3340	GGCGGGATAG	3350	ACTATGAAGA
3370	AGTGTTCGTC	3380	TTCCGCCAG	3390	TAAGCTATGT	3400	CTCCAGAAATG	3410	TCTTGTCAA	3420	TCAAGGGCGTT
3440	GGTCGCTTCC	3450	GGATTGTTA	3460	CATAACCGGA	3470	CATAATCATA	3480	CAAGGGTCA	3490	GCCTCTCTGA
3510	TTAACGCCCA	3520	GGGTTTCCC	3530	GGTATCCAGA	3540	TCCACAAACT	3550	TTAAAGACTT	3560	ACTTTACCGA
3580	CCGGCCCCGG	3590	TTTATCATCC	3600	CCCTCGGGTG	3610	TAATCAGAAT	3620	AAATGAAACA	3630	GGCCATATCC
3650	TTGTCCGTATC	3660	GGAAAGAGAT	3670	TGCAACCGGT	3680	TCCCCGACTT	3690	GTCTCAGTGA	3700	3570
3720	CCAGAAGCAA	3730	TTTCGTGTA	3740	ATTAGATAAA	3750	CAATCAGAGT	3760	CTTTGGCG	3770	3710
3790	ATAGGGTTGG	3800	TACTAGCAAC	3810	3880	3820	CTGAAGGGAT	3830	AGGTGGCCCC	3840	3780
3860	AAATCTATAC	3870	ATTAAGACGA	3880	CTCGAAATCC	3890	TATCCGAGTGT	3900	CGTAAAAACA	3910	3920
									TAGTAAACAT		TCCAAAACCG

FIG. 10D

60/75

3930	TGATGGAATG	3940	GAACAAACACT	3950	TAAAATCGCA	3960	GTATCCGGAA	3970	TGATTGATT	3980	GCCAAAAATA	3990	GGATCTCTGG	
4000	CATGCGAGAA	4010	4020	GCAGTTCTAT	4030	GGGGAAAGGGC	4040	4050	4060	TAGATCCAGA	4070	GGTAACCCAG	4080	4090
4070	GGATTGTTT	4140	TGTACCGATC	4150	AAAGGACTCT	4160	GGTACAAAAT	4170	CGTATTCACT	4180	AAAACGGGA	4190	GGTAGATGAG	
4140	ATGTGACGAA	4210	CGTGTACATC	4220	GAUTGAAATC	4230	CCTGGTAATC	4240	CGTTTTAGAA	4250	TCCATGATAA	4260	TAATTCTG	
4210	GATTGGT	4280	AATTTTTTTT	4290	GCACGTTCAA	4300	AATTTTTCG	4310	AACCCCTTT	4320	GGAAACAAA	4330	CACTACGGTA	
4280	GGCTGCGAAA	4350	TGTTCATACT	4360	GTTGAGCAAT	4370	TCACGTTCAT	4380	TATAATGTC	4390	GTTCGGGGC	4400	GCAAACGTGCAA	
4350	CTCCGATAAA	4420	TAACGGCCC	4430	AACACGGCA	4440	TAAAGAATTG	4450	AAGAGAGTTT	4460	TCACTGCATA	4470	CGACGATTCT	
4420	GTGATTGTA	4490	TTCAGCCCCAT	4500	ATCGTTTCA	4510	AGCTTCTGCC	4520	AACCGAACGG	4530	ACATTTCGAA	4540	GTATTGGCG	
4490	TACAGCCGG	4560	CCGTTAAC	4570	GGCGGGCTT	4580	CAATACCCCTG	4590	ATTGACTGGA	4600	ACAGCTGTAG	4610	CCCTGAACAG	
4560	CAGCGTGC	4630	TGCTGACGCG	4640	TCCGGCGATT	4650	TCCGCCTCTG	4660	CCGGACGGTC	4670	CCGGACGGTC	4680	AGCGATATT	
4630	TGGATAATGT	4700	AAAAAACGGC	4710	GGTGACGATG	4720	CCCTGCGGTGA	4730	ATACAGCGCT	4740	AAATTGATA	4750	AAACAGAAAGT	
4700	GACAGCGCTA	4770	CGCGTCACCC	4780	CTGAAGAGAT	4790	GGCGCCGCC	4800	GGCGCGGTCA	4810	4760	4830	ATTAAAACAG	
4770	GGGATGACCG	4840	CTGCCGTCAA	4850	AAATATTGAA	4860	ACGTTCCATT	4870	CCGCGCAGAC	4880	GCTACCCGCT	4890	GTAGATGTGG	
4840	AAACCCAGCC		AGGCAGGTGCGT		TGCCAGCAGG		TTACGCGTCC		CGTCTCGTCT		4900		ATATTCCGG	

FIG. 10E

61/75

4910	CGGCTGGCT	4920	CCGCCTTCT	4930	CAACGGTGT	4940	GATGCCGGC	4950	ACGCCGGCG	4960	GCATTGCCGG	4970
4980	GTGGTTCTGT	4990	GCTCGCCGCC	5000	GCCCATCGCT	5010	TCTATGCCGC	5020	GCAACTGTGT	5030	GGCGTGCAGG	5040
5050	AAATCTTTAA	5060	CGTCGGGGC	5070	GCGCAGGGGA	5080	TTTGCCGCTCT	5090	AGCGAGTCGG	5100	TACCGAAAGT	5110
5120	GGATAAAATT	5130	GCAACGCCCT	5140	TGTAAACCAGA	5150	GGCCTTCGGC	5160	AGGTCAAGCCA	5170	GCGTCTCGAC	5180
5190	GGCGCGGGCTA	5200	TCGATATGCC	5210	GGCGGGCGG	5220	TCTGAAGTAC	5230	AGACAGGGC	5240	GCAACACCGG	5250
5260	ATTTCCGTGGC	5270	TTCTGACCTG	5280	CTCTTCCAGG	5290	TGGTGATCGC	5300	AGACAGGGC	5310	TGCTGACGCC	5320
5330	TGATGCTGAC	5340	ATTGCCCGCA	5350	AGGTGGCGGA	5360	CCCGGATTCC	5370	CAGGTGATCC	5380	GCGGCCGAC	5390
5400	ACCGCCGGC	5410	AGGCCCTGAG	5420	CGCCAGTCGT	5430	CTGATTGTGA	5440	CGTAACGTGC	5450	GTCGCCATCT	5460
5470	CTAACATCGTA	5480	TGGGCCGGAA	5490	CACTTAATCA	5500	TCCAGACGGG	5510	AGGGCAGTGC	5520	ATGCCGATTAC	5530
5540	CAGCGCAGGC	5550	TCGGTATTTC	5560	TCGGCGACTG	5570	CAATGCCGGC	5580	GATTTGGTGG	5590	CGGAAACCAAC	5600
5610	CATGTTTAC	5620	CGACCTATGG	5630	CTATACTGCT	5640	ACCTGTTCCA	5650	ATTACGCTTC	5660	CAGAAACGGA	5670
5680	TGACCGTTCA	5690	GGAACTGTCTG	5700	AAAGCGGGCT	5710	GGCATCAACC	5720	AGGGATTTC	5730	TGGGGCGGGC	5740
5750	AGAACGGTCTG	5760	ACCGCCCAT	5770	AAAATGCCGT	5780	ATTGAAACAT	5790	ATTGAAACAT	5800	AGCATGAGCA	5810
5820	CTGAAAACAC	5830	TCTCAGCGTC	5840	GCTGACTTAG	5850	GTAAACGCC	5860	TCAAGGAGCA	5870	AGACATGGAT	5880

FIG. 10F

62/75

5890	5900	5910	5920	5930	5940	5950
AAGATACATT	GATGAGTTG	GACAAACCAC	AACTAGAAC	CAGTGAAAAA	AATGCTTAT	TTGTGAAATT
5960	5970	5980	5990	6000	6010	6020
TGTGATGCTA	TTGCCTTATT	TGTAACCATT	ATAAGCTGCA	ATAAAACAAGT	TAACAAACAC	AATTGCATTC
6030	6040	6050	6060	6070	6080	6090
ATTTTATGTT	TCAGGGTCAG	GGGGAGGGTGT	GGGAGGGTTT	TTAAAGCAAG	TAAAACCTCT	ACAAATGTGG
6100	6110	6120	6130	6140	6150	6160
TATGGCTGAT	TATGATCTCT	AGGGCCGCC	CTCGACGGCG	CGTCTAGAGC	AGTGTGGTTT	TCAAGAGGAA
6170	6180	6190	6200	6210	6220	6230
GCAAAAGGCC	TCTCCACCCA	GGCCTGGAAAT	GTTCACCCCC	AATGTCGAGC	AGTGTGGTTT	TGCAAGAGGA
6240	6250	6260	6270	6280	6290	6300
AGCAAAAGC	CTCTCCACCC	AGGCCTGGAA	TGTTTCCACC	CAAACCCCGC	CCAGCGTCTT	
6310	6320	6330	6340	6350	6360	6370
GTCATTGGCG	AATTGGAACCA	CCCATATGCA	GTCGGGGCGG	CGCGGTCCCA	GGTCCACTTC	
6380	6390	6400	6410	6420	6430	6440
GTGGCCGCTG	TGGCCTCGAA	CACCGAGCGA	CCCTGCAGCC	AATATGGGAT	CGGCCATTGA	ACAAGATGGA
6450	6460	6470	6480	6490	6500	6510
TTGCACGGCAG	GTTCTCCGGC	CGCTTGGGTG	GAGAGGCTAT	TCGGCTATGA	CTGGCACCAA	CAGACAATCG
6520	6530	6540	6550	6560	6570	6580
GCTGCTCTGA	TGCCGCCGTG	TTCCGGCTGT	TTCCGGGG	GGCCCCGGTT	CTTTTTGTCA	AGACCGACCT
6590	6600	6610	6620	6630	6640	6650
GTCCGGTGCC	CTGAATGAAC	TGCAGGTAAG	TGGGCCCGTC	GATGGCCGAG	GCGGCCCTCGG	CCTCTGCATA
6660	6670	6680	6690	6700	6710	6720
AATAAAAAAA	ATTAGTCAGC	CATGCATGGG	GCGGAGAAATG	GGCGGAACCTG	GGGGGAGTTA	GGGGGGGGAT
6730	6740	6750	6760	6770	6780	6790
GGCGGGAGTT	AGGGCGGGGA	CTATGGTTGC	TGACTAATTG	AGATGCATGC	TTTGCTACT	TCTGCCTGCT
6800	6810	6820	6830	6840	6850	6860
GGGGAGGCC	GGGACTTTCC	ACACCTGGTT	GCTGACTAAT	TGAGATGCAT	GCTTGGCATA	CTTCTGGCTG

FIG. 10G

63/75

6870	CTGGGAGCC	TGGGACTT	6890	CCACACCCTA	ACTGACACAC	6900	ATTCCACAGA	6910	ATTAATTCCC	6920	CTAGTTATTA
6940	ATAGTAATCA	ATTACGGGGT	6950	CATTAGTTCA	TAGCCCATAT	6970	ATGGAGTTCC	6980	GCGTTACATA	6990	ACTTACGGTA
7010	AATGGCCCGC	CTGGCTGACC	7020	GCCCCAACGAC	CCCCGCCCAT	7040	TGACGTCAAT	7050	AATGACCGTAT	7060	GTTCCCATAG
7080	TAACGCCAAT	AGGGACTTTC	7090	CATTGACGTC	AATGGGTGGA	7110	GTATTTCAGG	7120	TAAACTGCC	7130	ACTTGGCAGT
7150	ACATCAAAGTG	TATCATATGC	7160	CAAGTACGCC	CCCTATTGAC	7180	GTCAATGACG	7190	GTAATGGCC	7200	CGCCTGGCAT
7220	TATGCCCAGT	ACATGACCTT	7230	ATGGGACTTT	CCTACCTGCC	7240	AGTACATCTA	7250	CGTATTAGTC	7270	ATCGCTATTA
7290	CCATGGTGAT	GCGGTTTGG	7300	CAGTACATCA	ATGGCGTGG	7310	ATAGGGTTT	7320	GACTCACGGG	7340	GATTTCCAAG
7360	TCTCCACCCC	ATTGACGTCA	7370	ATGGGAGTTT	GTTTGGCAC	7380	CAAATCAAC	7390	GGGACTTTCC	7400	AAAATGTCGT
7430	AACAACCTCCG	CCCCATTGAC	7440	GCAAATGGGC	GGTAGGGCGTG	7450	TACGGTGGGA	7460	GGTCTATATA	7470	AGCAGAGGCTG
7500	GGTACCGTGA	CCGTCAGATC	7510	GCCTGGAGAC	TGCTCTGGCT	7520	GCCATCACAG	7530	7540	7550	GGGTCCCCGC
7570	TCAAGCTCTG	GGGCTCCTTC	7580	TGGGACAGA	AGGGGACAGA	7590	CCCAGGTGCC	7600	ATCTCTCACC	7610	CCAGTCTCCA
7640	TCTTCCCTGT	CTGCATCTGT	7650	AGGGGACAGA	GTACCATCA	7660	AGATGTGACA	7670	TCCAGATGAC	7680	AAGTCAGGAC
7710	ATTAAATTG	GTATCAGCAG	7720	AAACCAGGAA	AAGCTCCTAA	7730	CTTGCAGGCC	7740	7750	7760	ATTAGTTATT
7780	AAGTGGGTC	CCATCAAGGT	7790	TCAGCGGCAG	TGGATCTGG	7800	ACAGAGTTCA	7810	TATGTTGCAT	7820	CCAGTTTGCA
											7840 CAGCAGCCTG

FIG. 10H

64/75

7850	CAGCCTGAAG	7860	TTATTACTGT	7870	CTACAGGTT	7880	ATAGTACCCC	7890	TCGGACGTT	7900	GGCCAAGGGA
7920	CCAAGGGTGGAA	7930	ACATCAAACGT	7940	ACGGGTGGCTG	7950	CACCATCTGT	7960	CCGCCATCTG	7970	7980
7990	GAAATCTGGAA	8000	ACTGCCTCTG	8010	TTGTGTGCCT	8020	GCTGAATAAC	8030	TTCTATCCCA	8040	ATGAGGCAGTT
8060	AAGGTGGATA	8070	ACGCCCTCCA	8080	ATCGGGTAAC	8090	TCCCAGGGAGA	8100	GTGTACAGA	8110	AGTACAGTGG
8130	CCTACAGGCCT	8140	CAGCAGGCC	8150	GCAAGGCAGA	8160	CTACGAGAGC	8170	GCAGGACAGC	8180	AAGGACAGCA
8200	AGTCACCCAT	8210	CAGGGCCTGA	8220	GCTCGCCCGT	8230	CACAAAGAGC	8240	CTAACAGGG	8250	ACGCCTGCGA
8270	CGTTAACCGGT	8280	TACCAAAC TAC	8290	TTCGTGACAA	8300	CATGGGGCCG	8310	GAGAGTGTG	8260	8260
8340	CCTCGACTGT	8420	GCCTTCTAGT	8350	CTAGACTGGA	8360	TTCTTGGA	8370	TTGATATCTAC	8320	AATTCA GATC
8410	AGGTGCCACT	8490	CCCAC TGTCC	8430	TGCCAGCCAT	8440	CTGTTGTTG	8380	GTGACCTCCCT	8390	8330
8480	TCTATTCTGG	8560	GGGGTGGGGT	8500	TTTCCTAATA	8510	AAATGAGGAA	8450	TGACCCCTGGAA	8460	8400
8550	GGGATGCCGT	8630	GGGCTCTATG	8570	GGGGCAGGAC	8580	AGCAAGGGGG	8520	ATTGTCTGAG	8470	8470
8620	TAGGGGGGG	8690	ATGGGGGGAG	8640	GGACTATGGT	8650	AGGATTGGGA	8590	TGACCTGGCAT	8530	8540
8760	CTCTGCCTG	8770	TGGGGAGCC	8700	TTAGGGGGAG	8710	CAGCTGGGAC	8660	AGACAATAGC	8600	AGGCATGCTG
	TACTTCTGCC				TGGGGACTTT	8780	GCTGACTAAT	8670	TAGTGGCAAT	8670	TGGGGGAGT
					CCACACCTGG	8790	TTGCTGACTA	8730	TGAGATGCAT	8740	GCTTTGCATA
					TTCCACACCC		TAAC TGACAC	8800	ATTGAGATGC	8810	ATGCTTTGCA
									8820		8820
										GAATTAAATTC	

FIG. 10I

65/75

8830	CCCTAGTTAT	8840	TAATAGTAAT	8850	CAATTACGGG	8860	GTCATTAGTT	8870	CATAGCCCCAT	8880	CCGGGTTACA
8900	TAACCTACGG	8910	AAATGGCCC	8920	GCCTGGCTGA	8930	CGGCCCAACG	8940	ACCCCCGCC	8950	ATAATGACGT
8970	ATGTTCCCAT	8980	AGTAACGCCA	8990	ATAGGGACTT	9000	TCCATTGACG	9010	TCAATGGGTG	9020	GGTAAACTGC
9040	CCACTTGGCA	9050	GTACATCAAG	9060	GCCAAGTAGC	9070	GCCAAAGTAGC	9080	GAGTATTAC	9090	CGTAAATGG
9110	CCC GCCCTGGC	9120	ATTATGCCA	9130	GTACATGACC	9140	TTATGGGACT	9150	ACGTCAATGA	9160	9100
9180	TCATCGCTGT	9190	TACCATGGT	9200	GGCAGTACAT	9210	TTCCCTACTTG	9220	GCAGTACATC	9230	TACGTATTAG
9250	GGGATTCCA	9260	AGTCTCCACC	9270	ATGCCGTTTT	9280	CAATGGGGT	9290	GGATAGCGGT	9300	TTGACTCAGG
9320	CCAAATGTC	9330	GTAACAACTC	9340	CCATTGACGT	9350	CAATGGGAGT	9360	ACCAAATCA	9370	9310
9390	TAAGCAGAGC	9400	TGGGTACGTG	9410	CGCCCCATTG	9420	ACGCAAATGG	9430	TGTACGGTGG	9440	ACGGTCTATA
9460	TGCTCTTCCT	9470	TGTGGCTGTT	9480	AACCGTCAGA	9490	TCGCCCTGGAG	9500	CATGGTTGG	9450	9380
9530	GGCAAAGCCT	9540	GGGGGTCCC	9550	TGAGACTCTC	9560	ACGCCGTGGA	9570	GGGGAGTCTG	9460	AGCCTCATCT
9600	TACATGGACT	9610	GGGTCCGCCA	9620	GGCTCCAGGG	9630	TCCGGGTTCA	9640	GGTTCACCTT	9580	GGGGGGCTT
9670	ATCCCACATG	9680	GTACGGCAGAC	9690	CAGGGGCTGG	9700	AGTGGGTCTC	9710	GGTGGACTT	9650	9590
9740	GTTTCTCAA	9750	ATGAACAGGC	9760	TCCGTGAAGG	9770	CATCTCCAGA	9780	ACGTATTAGT	9720	CAATAACTAC
					TGAGAGCTGA		GGACACGGCT		GAGAACGCCA		AGTAGTGGTG
							GTCTATTACT		9790		9730
									GACTACAGGG		9800

FIG. 10J

66/75

9810	TCTGACTCCCT	GGGCCAGGG	9820	9830	ACGCCTGGTC	9840	CAGCTTCCT	9850	CAAGGGCCA	9860	TCGGTCTTCC
9880	CCCTGGCAC	CTCCCTCCAAG	9890	9900	AGCACCTCTG	9910	GGGCACAGC	9920	GGCCCTGGGC	9930	9940
9950	CCCCGAACCG	GTGACGGTGT	9960	9970	CGTGGAACTC	9980	AGGGCCCTG	9990	TGCACACCTT	10000	AGGACTACTT
10020	CTACAGTCCT	CAGGACTCTA	10030	10040	CTCCCTCAGC	10050	ACCAGGGCG	10060	CCGGGCTGTC	10070	10010
10090	CCTACATCTG	10100	10110	10120	GCAAGCCC	10130	GGGGACAAG	10140	GGCACCCAGA	10080	10080
10160	TGACAAAACT	CAACGTGAAT	10170	10180	AGCACCTGAA	10190	CTCCTGGGG	10200	CAGCAGCTTG	10140	10150
10230	CCCCAAAAC	CACACATGCC	10240	10250	CCTCATGATC	10260	CTCGGGACCC	10270	AAAGTTGAGC	10070	10220
10300	GCCACGAAGA	CCCTGAGGTC	10310	10320	AAGTTCAACT	10330	CTGAGGTGAC	10340	CTTCCCTCTTC	10290	10290
10370	GCGGGGGAG	GAGCAGTACA	10380	10390	ACAGCACGTA	10400	GGGCGTGGAG	10410	GTGGACGTGA	10210	10210
10440	CTGAATGGCA	AGGAGTACAA	10450	10460	GTGCAAGGTC	10470	AGCGTCCTCA	10480	CCAAGACAAA	10130	10130
10510	CCAAAGCCAA	AGGGCAGCCC	10520	10530	CGAGAACAC	10540	CCCTCCCAAG	10550	GTGGACTGG	10070	10070
10580	GAACCAGGTC	AGCCTGACCT	10590	10600	GCCTGGTCAA	10610	AGGCTTCTAT	10620	TGCGGTGGA	10630	10640
10650	AATGGGCAGC	10660	10670	10680	CTACAAGACC	10690	CCAGCGACA	10700	GGCAGCAGGG	10710	10710
10720	ACAGCAAGCT	10730	10740	10750	AAGAGCAGGT	10760	ACGCCCTCCCG	10770	TTCTTCCTCT	10780	10780

FIG. 10K

FIG. 10L

10790	GGCTCTGCAC	AACCACTACA	CGCAGAAAGAG	10810	CCTCTCCCTG	TCTCCGGGT	10830	AATGAGGATC	CGTTAACGGT	10850
10860	TACCAACTAC	CTAGACTGGA	10880	TTCGTGACAA	10890	CATGCCGGCG	10900	TGATATCTAC	CTCGACTGT	10920
10860	TACCAACTAC	CTAGACTGGA	10880	TTCGTGACAA	10890	CATGCCGGCG	10900	TGATATCTAC	CTCGACTGT	10920
10930	GCCTTCTAGT	TGCCAGCCAT	10940	CTGTTGTTGC	10950	CCCCCCCC	10960	TGACCCCTGGA	AGGTGCCACT	10990
11000	CCCACTGTCC	TTTCTTAATA	11010	AAATGAGGAA	11020	ATTGCATCGC	11030	TAGGTCTGAG	TCTATTCTGG	11060
11070	GGGGTGGGT	GGGCAGGGAC	11080	AGCAAGGGGG	11090	AGGATTGGGA	11100	ATTGTCTCAT	TCTATTCTGG	11060
11140	GGGCTCTATG	GCTTCTGAGG	11150	CGGAAAAGAAC	11160	CAGCTGGGC	11170	AGACAATAGC	GGGATGCCGT	11130
11210	TAACTCTCTC	CTCCCTCCTT	11220	TTCCCTGCAG	11230	GACGAGGCAG	11240	AGGCATGCTG	GGGATGCCGT	11130
11280	TTCCCTGCC	AGCTGTGCTC	11290	GACGTTGTCA	11300	CTGAAGCGGG	11310	CGCTAGGCTCG	GGGATGCCGT	11130
11350	GGGGCAGGAT	CTCCTGTCAT	11360	CTCACCTTGC	11370	TCCTGCCGAG	11380	AAAGTATCCA	GGGATGCCGT	11130
11420	CGGCTGCATA	CGCTTGATCC	11430	GGCTACCTGC	11440	CCATTGACC	11450	TCATGGCTGA	GGGATGCCGT	11130
11490	GTACTCGGAT	GGAAAGCCGGT	11500	CTTGTGCGATC	11510	AGGATGATCT	11520	CTGCTATTGG	GGGATGCCGT	11130
11560	CGAACTGTTTC	GCCAGGTAAG	11570	TGAGCTCCAA	11580	TTCAAGCTCT	11590	CGAGCTAGGG	GGGATGCCGT	11130
11630	TTCTCAATT	CTTATTTGCA	11640	TAATGAGAAA	11650	AAAAGAAAAA	11660	TTAATTAA	GGGATGCCGT	11130
11700	GAGCAAATGC	GTTGCCAAA	11710	AGGATGCCTT	11720	11730	11740	TTCTCTGCAC	AGATAAGGAC	11130

67/75

68/75

11770	CAGAGGGAGT	11780	ACCCAGAGCT	11790	GAGACTCCTA	11800	TGCCAGTGAG	11810	TCCAGGGAGA	11820	AATATGCTTG
11840	TCATCACCGA	11850	AGCCTGATTTC	11860	CGTAGAGCCA	11870	AGGCCATTGA	11880	TGCTCACACA	11890	GGATAGAGAG
11910	GGCAGGAGCC	11920	AGGCAGAGC	11930	ATATAAGGTG	11940	AGTTAGGATC	11950	CACATTTGCT	11960	TCTGACATAG
11980	TTGTGTTGGG	12000	AGCTTGGATA	12000	GCTTGGGGGG	12010	GGGACAGCTC	12020	AGGGCCTGGGA	12030	AACTTGACGG
12050	CAATCCCTAGC	12060	GTGAAGGGCTG	12070	GTAGGATTTT	12080	ATCCCCGCTG	12090	CCATCATGGT	12100	AACTGCATCG
12120	TCGCCGTGTC	12130	CCAAAATATG	12140	GGGATTGGCA	12150	CCTACCCCTGG	12160	CGGACCATTG	12170	12180
12190	CAAGTACTTC	12200	CAAAGAATGA	12210	CCACAACCTC	12220	GGTAAACAGA	12230	CCTCCGCTCA	12240	GGAAACGAGT
12260	AAAACCTGGT	12270	TCTCCATTCC	12280	TGAGAACGAT	12290	TTCAGTGGAA	12300	ATCTGGTGT	12310	TATGGTAGGG
12330	AACTCAAAGA	12340	ACCAACCAGA	12350	GGAGCTCATT	12360	CGACCTTTAA	12370	TAATATAGTT	12380	12390
12400	CCATTAAGAC	12410	TTATTGAACA	12420	ACCGGAATTG	12430	AGGACAGAAAT	12440	GATGCCTTAA	12450	CGTAGGGCGG
12470	CTGTTACCA	12480	GGAAAGCCATG	12490	AATCAACCAG	12500	TAGACATGGT	12510	TTGGATAGTC	12520	12530
12540	TGAAAGTGAC	12550	ACGTTTTTCC	12560	CAGAAATTGA	12570	ACTCTTGTG	12580	ACAAGGATCA	12590	GGAGGGCAGTT
12610	CTCTCTGAGG	12620	TCAAGGAGGA	12630	AAAAGGCATC	12640	TATAAACCTTC	12650	TCCCAGAATA	12660	12670
12680	AAGATGCTTT		CAAGTCTCT	12700	GCTCCCTCC	12710	AAGTATAAGT	12720	TTGAAGTCTA	12730	GACTAACAGG
							TAAAGCTATG		CATTTTTATA		12740
									AGACCATGGG		ACTTTTGCTG

FIG. 10M

69/75

12750	GCTTTAGATC	12760	AGCCTCGACT	12770	GTTGCCCTCTA	12780	ATCTGCCAGCC	12790	TGCCCTTGT	12800	CCGTGCCTTC
12820	CTTGACCCCTG	12830	GAAGGGTCCA	12840	CTCCCCACTGT	12850	CCTTTCCCTAA	12860	AAATTGAGG	12870	GCATTGCTG
12890	AGTAGGTGTC	12900	ATTCTATTCT	12910	GGGGGGTGGG	12920	GTGGGGCAGG	12930	ACAGCAAGGG	12940	GAAGACAAATA
12960	GCAGGCATGC	12970	TGGGGATGCG	12980	GTGGGCTCTA	12990	GGGGAAAAGA	13000	ACCAGCTGGG	13020	GCTCGAAGCG
13030	GCGGCCATT	13040	TCGCTGGTG	13050	TCAAGATGCCG	13060	GATGGCGTGG	13070	13080	13090	ACTGAGGTTT
13100	TCCGCCAGAC	13110	GCCACTGCTG	13120	CCAGGGGCTG	13130	ATGTGCCCGG	13140	GACGGGGGG	13150	TTCCGGTTGCA
13170	CTACCGTAC	13180	TGTGAGCCAG	13190	AGTTGCCCGG	13200	CTTCTGACCA	13210	TGCGGTGCG	13220	CAATCAAATG
13240	TTTACCTTGT	13250	GGAGCGACAT	13260	CCAGAGGCAC	13270	CGCTCTCCGG	13280	TGAGGTTAGT	13290	13300
13310	CAGTGCAGGA	13320	GCTCGTTATC	13330	GCTATGACGG	13340	TTCACCGCTT	13350	TACCATCCAG	13360	CGCCACCATC
13380	AACGGAACTG	13390	GAAAAACTGC	13400	TTGCTGGTGT	13410	CGCTGGTCAC	13420	TGCGATGGTT	13440	TGCCCCGATA
13450	GACCAGACCG	13460	TTCATACAGA	13470	ACTGGGGATC	13480	CGCGCTGGTA	13490	GGGGGGTGC	13500	GGGGGGCAA
13520	GGGTTGCCGT	13530	TTTCATCATA	13540	TTAACATCAGC	13550	TCGCCAAAT	13560	13570	13580	AGCCGACAC
13590	GGGGATACTG	13600	ATCAGGAAAGG	13610	TTAACATCAGC	13620	CACCGCCGTA	13630	GACGAAGCCG	13640	CCCTGTAAAC
13660	ATCAGGGGGC	13670	ACGAAACGCC	13680	TGCCAGTATT	13690	CCCAAGAAC	13700	TTACCCATCG	13720	CGTGGGGGTA
	TTCGCAAAGG				GGGTCTCTCC		AGGCATTTTT		13710		TTTCGGCACCA

FIG. 10N

70/75

13730	GCCGGAAAGG	13740	ATCGGTCTTC	13750	ATCCACGCGC	13760	GGCAAATAAT	13770	ATCGGGTGGCC	13780	GTGGGTGTCGG
13800	CTCCGCCGCC	13810	TTCATACTGC	13820	ACCGGGCGGG	13830	AGATTGATCGAC	13840	CAGCGATACA	13850	GGCGGTCGTG
13870	ATTAGGCCCG	13880	TGGCCTGATT	13890	CAITCCCCAG	13900	CGACCAGATG	13910	GGTGATTACG	13920	ATCGGGCTGC
13940	ACCATTGCGC	13950	TTACGGCTTC	13960	GGTAGGCCAGC	13970	GGGGATCATC	13980	GGTCAGACGA	13990	TTCATTGGCA
14010	CCATGCCGTG	14020	GGTTTCAATA	14030	GCTCATGCC	14040	GGGTGGCTAG	14050	CGGTGGCAC	14060	GGGTGTACCA
14080	CAGCGGATGG	14090	TTCGGATAAT	14100	TTGGCTTCAT	14110	CCACCACATA	14120	GCTTCATCAG	14130	CAGGATATCC
14150	TGCACCATCG	14160	TCTGCTCATC	14170	GGGAACAGCG	14180	AAGTTGTTCT	14190	GCTTCATCAG	14200	ACGCCCTCGAA
14220	TCAGCAACGG	14230	CTTGGCGTTTC	14240	CATGACCTGA	14250	GATGATGCTC	14260	GTGACGGTTA	14270	14280
14290	GGCTTCTGCT	14300	TCAATCAGCG	14310	AGCAGCAGCA	14320	AATCCGCACC	14330	14340	14350	CGACATCGCA
14360	ACAGTTTCGG	14370	TGCCGTGGC	14380	GGTGTGCAGT	14390	TCAACCACCG	14400	CACGATAGAG	14410	ATTGGGATT
14430	CATCGATAAT	14440	TTCACCGGG	14450	TTCAGACGTA	14460	GTGTGACGGC	14470	14480	14490	CCACCAAGCT
14500	TGTTACCCGT	14510	AGGTAGTCAC	14520	TGCCGCTGGC	14530	GACCTGCCTT	14540	TCACCCCTGCC	14550	ATAAAAGAAC
14570	TCATCATTAA	14640	CACGGAAAT	14650	AAAGGGCGGG	14660	ACTTCAGCCT	14610	CCAGTACAGC	14620	GGGGCTGAAA
					GCACATCTGA		TGCTGATTTT				AACGAGACGT
					14590	14600	GTGTAGTCGG	14680	TTTATGCAGC	14690	14700
					14660	14670	CAGATAACTG		CCGTCACCTCC		AGCGCAGCAC

FIG. 10P

71/75

14710	CATCACCGCG	14720	AGGGGGTTT	14730	CTCCGGCGG	14740	TAAAATGCG	14750	CTCAGGTCAA	14760	14770
14780	TCCTGGCGT	14790	AACCGACCCAA	14800	GCGCCCCGTTG	14810	CACCACAGAT	14820	GAAACGCCGA	14830	14840
14850	TTGCGCTTG	14860	GCCTCCTGT	14870	AGCCAGGCTTT	14880	CATCAACATT	14890	AAATGTGAGC	14900	TCAAAAATAA
14920	CTCCGTGGGAA	14930	ACAAACGGCG	14940	GATTGACCGT	14950	AATGGGATAG	14960	GAGTAACAAC	14970	14980
14990	CGGTGCATCT	15000	GCCAGTTGA	15010	GGGGACGGACG	15020	ACAGTATCGG	15030	GTCACGTTGG	15040	CGCATCGTAA
15060	CGGCACCGC	15070	TTCTGGTGCC	15080	GGAAACCAGG	15090	CCTCAGGAAG	15100	ATCGGACTCC	15110	15050
15130	AGGGGATCG	15140	GTGCGGGCCT	15150	GCAAGCGCCA	15160	TTGCCATTTC	15170	AGGCTGGCA	15180	AGCCAGGCTT
15200	AGTTGGTAA	15210	GGCCAGGGTT	15220	CTTCGCTATT	15230	ACGCCAGCTG	15240	GATGTGCTGC	15250	AAGGGGATTA
15270	GAAGCAGACG	15280	ACCTTCCGTT	15290	TTCCCAGTCA	15300	CGACGTTGTA	15310	ATCCGTTGAG	15320	GGGCTGCCCT
15340	TAAACCAGAA	15350	GTGCAGCCAG	15360	CGGCAGCTGC	15370	GCCCCGCTGC	15380	ACAATCGTGC	15390	15260
15410	CCACCAAC	15420	CAAATTATAC	15430	GGGGGCCACC	15440	CCTTCTCCCG	15450	TGCCTAACAT	15460	15330
15480	CTGCACCTT	15550	CACCAACATC	15560	GATGTCTGAA	15570	CTCCACCAAT	15580	GGGACGGAA	15590	GGAAACAAAC
15620	ACGGTCAAA	15630	AGACGACAGA	15640	CAACAATTGT	15650	TGGAAGCTAT	15660	AAAAATCGCA	15670	15680
	ACATTGAGC		CTAAACCGCC		CGCCCCGAAAC		CAGTACAATA		GTAAGGTGTC		CCAGCCCCGC

FIG. 10Q

FIG. 10R

15690	CTGCCCTCC	ACCGATGGT	15700	GATTATCAT	15710	CAGCTCCACC	15720	ACCGCCGCCA	15730	TTAGTAGATT	15740	TGCCGTCTGA	15750	
15760	AATGTTACCA	CCGCCTGCAC	15770	CATCGCTTTC	15780	TAACGTGTTG	15790	TCTGAATTAA	15800	ATCGGGCAC	15810	AGTTAGATTG	15820	
15830	AAACCCGCC	GCAATCAGAA	15840	ATAATTCCAA	15850	AAGCTCAAC	15860	TACAAATTG	15870	ATCGGGACG	15880	ATCGGGACG	15890	
15900	TGTTAGCCGA	CACAAATTAA	15910	AGCGCTCGTG	15920	TGGCTATGGC	15930	AAATCGTCT	15940	TCGGAAGCAA	15950	CTTCTAACGA	15960	
15970	CGAGGGTTGG	GACGACGACG	15980	ATAATCGGCC	16000	TAATAAGCT	16010	AAACACGCCG	16020	ATGTTAAATA	16030	TGTCCAAGCT	16030	
16040	ACTAGTGGTA	CCGCTTGGCA	16050	GAACATATCC	16060	ATCGCGTCCG	16070	CCATCTCCAG	16080	CAGCCGCACG	16090	CGGCCATCT	16100	
16110	CGGGCAGCGT	TGGGTCTGG	16120	CCACGGGTGC	16130	GCATGATCGT	16140	GCTCCTGTCG	16150	TTGAGGACCC	16160	GGCTAGGCTG	16170	
16180	GGGGGGTTGC	CTTACTGGTT	16190	16200	ATCACCGATA	16210	CGCGAGCCAA	16220	CGTGAAGCGA	16230	CTGCTGCTGC	16240	CTGAAACGCG	16240
16250	AAAACGTCTG	CGACCTGAGC	16260	AGCAGAAATGA	16270	ATGGTCTTCG	16280	GGTTCGGTGT	16290	TTCGTAAAAGT	16300	CCTGTGGAAC	16310	
16320	GGAAAGTCAGC	GCCCTGCACC	16330	AAACAACATGA	16340	ATGTTGCAT	16350	CGAGGATGC	16360	CGAGGATGC	16370	CGAGGATGC	16380	
16390	ACCTACATCT	GTATTAACGA	16400	AGCGCTGGCA	16410	GGATCTGCAT	16420	GTGATTTC	16430	TCTGGTCCC	16440	CCGCATCCAT	16450	
16460	ACCGCCAGTT	GTTTACCCCTC	16470	ACAACGTTCC	16480	TTGACCCCTGA	16490	GTGATTTC	16500	ATCAGTAACC	16510	CGTATCGTGA	16520	
16530	GCATCCTCTC	TCGTTTCATC	16610	GGTATCATTAA	16620	AGTAACCGGG	16550	CATGTTCATC	16560	CAGAAATCCC	16570	GCTTACACGG	16590	
16600	GACCAAAACAG	GAAAAAACCG		CCCTTAACAT		CCCCCATGAA	16630	ATCAGAAGCC	16640	AGACATTAAC	16650	GCTTCTGGAG	16660	

72/75

73/75

16670	AAACTCAACG	16680	AGCTGGACGC	16690	GGATGAACAG	16700	GCAGACATCT	16710	GTGAATCGCT	16720	TCACGACCAC	16730
16740	TTTACCGCAG	16750	CTGCCTCGCG	16760	CGTTTCGGTG	16770	ATGACGGGTGA	16780	AAACCTCTGA	16790	TCCCGGAGAC	16800
16810	GGTCACAGCT	16820	TGTCTGTAAG	16830	CGGATGCCGG	16840	GAGCAGACAA	16850	GCCC GTCA GG	16860	GGCGGT CAGC	16870
16880	GGGTGTCGGG	16890	GCGCAGCCAT	16900	GACCCAGTCA	16910	CGTAGCGATA	16920	GCGGAGTGT A	16930	GCGCGT CAGC	16940
16950	ATCAGAGCAG	16960	ATTGTACTGA	16970	GAGTGACCCA	16980	TATGGGGTGT	16990	GAAAATACCGC	17000	ACAGATGC GT	17010
17020	TACCGCATCA	17030	GGCGCTCTTC	17040	CGCTTCCCTG	17050	CTCAC TGACT	17060	GGCTGGC GCTC	17070	CTGC GGCG AG	17080
17090	CGGTATCAGC	17100	TCACTCAAAG	17110	GGG GTAATAC	17120	GGTTATCCAC	17130	AGAATCAGGG	17140	GAAAGAACAT	17150
17160	GTGAGC AAA	17170	GGCCAGCAA	17180	AGGCCAGGAA	17190	CCG TAAAAG	17200	GCCGGTTGC	17210	CCATAGGCTC	17220
17230	CGCCCCCTG	17240	ACGAGGCATCA	17250	CAAAAATCGA	17260	CGCTCAAAGTC	17270	AGAGGTGGCG	17280	GGACTATAAA	17290
17300	GATA CAGGC	17310	GTTC CCCCCT	17320	GGAAAGCTCCC	17330	TCGTGGCCTC	17340	ACCCCTGCCG	17350	TTACCGGATA	17360
17370	CCTGTCCGCC	17380	TTCTCCCTT	17390	CGGGAAAGCGT	17400	GGCGCTTTCT	17410	GCTGTAGGTA	17420	TCTCAGTTCG	17430
17440	GTGTAGGT CG	17450	TTCGCTCCAA	17460	GCTGGGCTGT	17470	GTGCACGAAC	17480	CCCCGTTCA	17490	TGGCCTTAT	17500
17510	CGGGTA ACTA	17520	17530	17540	TAAGACACGA	17550	TGCTACAGAG	17560	GGCGACCGC	17570	CCACTGGTAA	17580
17580	CAGGATTAGC	17590	TCGTCTTGAG	17600	TC CAACCCGG	17610	TTCTTGAAGT	17620	CTTATCGCCA	17630	CTGGCAGCAG	17640

FIG. 10S

74/75

17650	ACTAGAAGGA	17660	CAGTATTGG	17670	TATCTGGC	17680	CTGCTGAAGC	17690	CAGTTACCTT	17700	CGAAAAAGA	17710
17720	CTTGATCCGG	17730	CAAACAAACC	17740	ACCGCTGGTA	17750	GGGGTGGTT	17760	TTTGTGTC	17770	AAGCAGCAGA	17780
17790	AAAAAAAGGA	17800	TCTCAAGAAG	17810	ATCCCTTGAT	17820	CTTTCTACG	17830	GGGTCTGACG	17840	CTCAGTGGAA	17850
17860	CGTTAAGGGA	17870	GAGATTATCA	17880	AAAAGGATCT	17890	TCACCTAGAT	17900	CAATGCTTAC	17910	TAAAATGAA	17920
17930	GTTTAAATC	17940	ATATATGAGT	17950	AAACTGGTC	17960	TCACCTAGAT	17970	CCCTTAAAT	17980	TCAGTGGGC	17990
18000	ACCTATCTCA	18010	GGGATCTGTC	18020	ATCCATAGTT	18030	TGACAGTTC	18040	CCGTGACTCC	18050	GATAACTACG	18060
18070	ATACGGGAGG	18080	GCTTACCATC	18090	TGGCCCCAGT	18100	GCTGCAATGA	18110	CCCACGCTCA	18120	CCGGCTCCAG	18130
18140	ATTATATCAGC	18150	AATAAACCAG	18160	CCAGCCCCAA	18170	GGGGCGAGGC	18180	CCACGCTCA	18190	TATCCGCCCTC	18200
18210	CATCCAGTCT	18220	ATTAAATTGTT	18230	GGCGGGAAAGC	18240	CAGAAAGTGGT	18250	CCTGCAACTT	18260	GGCAACGTT	18270
18280	GTGCCCCATTG	18360	CTGCAGGCAT	18300	TAGAGTAAGT	18310	AGTTGCCAG	18320	TTAATAGTTT	18330	TCCGGTTCCC	18340
18350	AACGATCAAG	18430	GGGAGTTACA	18370	GGCTCGTGTCA	18380	TTGGTATGGC	18390	AGCTCCTTCG	18400	GTCCTCCGAT	18410
18420	CGTTGTCAGA	18490	AGTAAGTTGG	18510	TGATCCCCA	18520	AAAAGCGGTT	18460	CACTGCATAA	18470	TTCTCTTACT	18480
18560	GTCATGCCAT	18570	CCGTAAGATG	18580	CCGCAGTGT	18590	ATCAACCAA	18600	GTCAATTCTGA	18610	GAATAGTGTA	18620
	TGGGGGACCC		GAGTTGCTCT		TGCCCGGGGT		CAACACGGGA		TAATACCGCG		GAACATAGCA	

FIG. 10T

75/75

18630	AGTGCTCATC	18640	ATTGGAAAC	18650	GTTCTTCGGG	18660	GCGAAAACTC	18670	TCAAGGATCT	18680	GAGATCCAGT
18700	TCGATGTAAC	18710	CCACTCGTC	18720	ACCCAACGTGA	18730	TCTTCAGGCAT	18740	CTTTTACTTT	18750	TCTGGGTGAG
18770	CAAAACAGG	18780	AAGGCCAAAT	18790	GCCGCCAAAAA	18800	AGGGATAAAG	18810	GGGACACGG	18820	TACTCATACT
18840	CTTCCTTTT	18850	CAATATTATT	18860	GAAGCATTAA	18870	TCAGGGTTAT	18880	TGTCCTCATGA	18890	GCGGATACAT
18910	ATTTAGAAA	18920	ATAAACAAT	18930	18940	18950	18960	18970	18980	18990	ATTTGAATGT
18980	CCATTATTAT	19050	CATGACATTA	19060	ACCTATAAAA	19070	ATAGGGTAT	19080	CACGAGGCC	19090	GTCTAAGAAA

FIG. 10U

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 98/03935

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/00	C12N15/85	C12Q1/68	C12N5/10	C12N9/12
C12N15/13	C07K16/28	C12N15/12	C07K14/705	G01N33/53
C12N15/62	C07K19/00			

According to International Patent Classification(IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C12Q C07K G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 94 11523 A (IDEC PHARMACEUTICALS CORPORATION (US); REFF MITCHELL E. (US)) 26 May 1994 cited in the application see abstract see page 9, line 21 - page 10, line 29 see page 41, line 19 - page 42, line 19; figure 6 ---	1, 4-8, 11, 12, 25-29, 31, 32
A	US 5 464 764 A (CAPECCHI MARIO R. AND KIRK THOMAS R.) 7 November 1995 see abstract see column 13, line 32 - column 14, line 5 ---	1
A	WO 94 05784 A (UNITED STATES AMERICA REPRESENTED BY THE SECRETARY US DPT. AGRICULTURE) 17 March 1994 see abstract ---	1
	-/-	

 Further documents are listed in the continuation of box C. Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
23 July 1998	05/08/1998
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Macchia, G

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 98/03935

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 93 24642 A (TSI CORPORATION (US)) 9 December 1993 see abstract ----	1
A	BARNETT R.S. ET AL.: "Antibody production in chinese hamster ovary cells using an impaired selectable marker" ACS SYMPOSIUM SERIES: ANTIBODY EXPRESSION AND ENGINEERING, vol. 604, 1995, pages 27-40, XP002072464 -----	

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 98/03935

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9411523	A 26-05-1994	AU AU CA DE EP ES JP US US	682481 B 5613294 A 2149326 A 669986 T 0669986 A 2088838 T 8503138 T 5648267 A 5733779 A	09-10-1997 08-06-1994 26-05-1994 10-10-1996 06-09-1995 01-10-1996 09-04-1996 15-07-1997 31-03-1998
US 5464764	A 07-11-1995	US US US	5487992 A 5627059 A 5631153 A	30-01-1996 06-05-1997 20-05-1997
WO 9405784	A 17-03-1994	AU MX	4839493 A 9305183 A	29-03-1994 31-05-1994
WO 9324642	A 09-12-1993	AU	4401993 A	30-12-1993